

# A Justification of Conditional Confidence Intervals

Citation for published version (APA):

Beutner, E., Heinemann, A., & Smeeke, S. (2017). *A Justification of Conditional Confidence Intervals*. arXiv.org No. 1710.00643

## Document status and date:

Published: 02/10/2017

## Document Version:

Early version submitted to journal

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

# A Justification of Conditional Confidence Intervals

Eric Beutner<sup>†</sup>

Alexander Heinemann<sup>†</sup>

Stephan Smeekes<sup>†</sup>

<sup>†</sup>Department of Quantitative Economics

Maastricht University

October 3, 2017

## Abstract

To quantify uncertainty around point estimates of conditional objects such as conditional means or variances, parameter uncertainty has to be taken into account. Attempts to incorporate parameter uncertainty are typically based on the unrealistic assumption of observing two independent processes, where one is used for parameter estimation, and the other for conditioning upon. Such unrealistic foundation raises the question whether these intervals are theoretically justified in a realistic setting. This paper presents an asymptotic justification for this type of intervals that does not require such an unrealistic assumption, but relies on a sample-split approach instead. By showing that our sample-split intervals coincide asymptotically with the standard intervals, we provide a novel, and realistic, justification for confidence intervals of conditional objects. The analysis is carried out for a general class of Markov chains nesting various time series models.

**Key words:** Conditional confidence intervals, Parameter uncertainty, Markov chain, Sample-splitting, Prediction, Merging

**JEL codes:** C53, C22, C32, G17

# 1 Introduction

One of the big open questions in time series is how to quantify uncertainty around point estimates of conditional objects such as conditional means or conditional variances. A fundamental issue arises in the construction of confidence intervals that ought to capture the parameter estimation uncertainty contained in these objects. This fundamental issue stems from the fact that on one hand one must *condition* on the sample as the past informs about the present, yet on the other hand one must allow the data up to now to be treated as *random* to account for estimation uncertainty. In practice it is typically ignored and confidence intervals are commonly, yet inappropriately, constructed by treating the sample *simultaneously* as fixed and random. Frequently, such approach is motivated by pretending to have two independent processes. Assuming two independent processes with the same stochastic structure, using one for conditioning and one for the estimation of the parameters, bypasses the issue. However, the central assumption turns out to be not realistic and violated in nearly all circumstances; with the exception of perhaps some experimental settings, it is hard to think of any application where a researcher has a replicate, independent of the original series, at hand. As such, the theoretical justification for the intervals commonly constructed by practitioners is flimsy at best. Therefore it is the objective of the present paper to develop an alternative, realistic justification for such confidence intervals around point estimates of conditional objects.

In the literature the fundamental issue described above is encountered in various ways. In the very specific case of a first-order autoregressive (AR) process with Gaussian innovations, Phillips (1979) investigates the statistical dependence between the ordinary least squares (OLS) estimator and the endogenous variable conditioned upon. He obtains an Edgeworth-type expansion for the law of the conditional mean

and, further, studies forecasting, where the fundamental issue equally arises.<sup>1</sup> It is also recognized by Schmidt (1977) when forecasting a structural model, who declares it to be “conceptually a problem” (p. 998). More recent studies investigate parameter uncertainty by using resampling methods, that typically mimic a distribution in which the sample, or at least a sub-sample, is treated as fixed and random at the same time (c.f. Pascual et al., 2004, 2006, Pan and Politis, 2016a, 2016b). Aware of this paradox, Kreiss (2016) points out that conditioning on observing specific *in-sample* values affects the parameter estimator, but the effect is often erroneously disregarded. Deviating from the various bootstrap approaches, Hansen (2006) examines parameter uncertainty in interval forecasts in a classical statistical framework. Similar to general regression, yet inappropriate in dynamic models, he conditions on an arbitrary fixed *out-of-sample* value to avoid the fundamental issue. In contrast, Lütkepohl (2005, p. 95) explicitly states a two-independent-processes assumption in connection with vector AR models. He postulates that such assumption is asymptotically equivalent to using only data not conditioned upon for estimation and refers to Samaranayake and Hasza (1988). The latter show that the difference between the conditional mean and its estimated counterpart vanishes almost surely. Acknowledging the fundamental issue while avoiding the two independent processes argument bears careful statements as in Francq and Zakoïan (2015) who write in view of this issue “the delta method ... suggests” (p. 162). In line Pesaran (2015) recognizes that although such intervals “have been discussed in the econometrics literature, the particular assumptions that underlie them are not fully recognized” (p. 389).

This paper provides a novel, and realistic, justification for commonly constructed confidence intervals around point estimates of conditional objects. Our solution is based on a simple sample-split approach and a weak dependence condition, which

---

<sup>1</sup>For prediction intervals some solutions have been discussed. We refer to Section 5.

allows to partition our sample into two asymptotically independent subsamples. For a rich class of Markovian models we construct asymptotically valid sample-split intervals, without relying on the assumption of observing two independent processes, and show that these intervals coincide asymptotically with the intervals commonly constructed by practitioners. To the best of our knowledge, except for Belyaev and Sjöstedt-De Luna (2000), this paper is the only one to study merging, a concept generalizing the notion of weak convergence, in the context of conditional distributions. Moreover, our paper seems to be the first to employ merging of conditional distributions in time series. By employing this concept we avoid unrealistic assumptions such as observing  $X_n = x$  (in dynamic models), losing the time index, and instead explicitly acknowledge that the conditional objects vary over time.

The rest of the paper is organized as follows: Section 2 illustrates the issue at hand by means of two simple examples. Section 3 specifies the general class of Markov chains and describes the argument of two independent processes as well as our sample-split approach. In Section 4 we establish merging among the proposed and the two-independent-processes estimator in probability under mild conditions. Further, we construct asymptotically valid sample-split intervals and show that these coincide asymptotically with the standard intervals. The extension to prediction is discussed in Section 5. Section 6 concludes. Proofs of the main results are collected in Appendix A, whereas Appendix B contains additional proofs.

## 2 Examples

For illustration of the fundamental issue at hand, we consider two simple examples. Let  $\gamma \in (0, 1)$  and denote the sample of a time series  $\{X_t\}$  of length  $n \in \mathbb{N}$  by  $\mathbf{X}_n = (X_1, \dots, X_n)'$  with corresponding sample path  $\mathbf{x}_n = (x_1, \dots, x_n)'$ . We discriminate

between random variables and their realized counterparts by writing the former in capital and the latter in lowercase letters to avoid ambiguity. First, we consider the conditional mean in a first-order AR model.

**Example 1.** Suppose the time series  $\{X_t\}$  follows an AR(1) process given by

$$X_t = \beta X_{t-1} + \varepsilon_t, \quad (2.1)$$

where  $|\beta| < 1$  and  $\{\varepsilon_t\}$  are independent and identically distributed (*i.i.d.*) with  $\mathbb{E}[\varepsilon_t] = 0$  and  $\mathbb{E}[\varepsilon_t^4] < \infty$ . The goal is to provide a  $100(1 - \gamma)\%$  confidence interval for

$$\mathbb{E}[X_{n+1}|X_n = x_n] = \beta x_n, \quad (2.2)$$

the conditional mean of the one step ahead observation. A natural estimate for (2.2) is given by  $\hat{\beta}(\mathbf{x}_n)x_n$ , where  $\hat{\beta}(\mathbf{x}_n)$  denotes the OLS estimate of  $\beta$  depending on the sample realization  $\mathbf{x}_n$ . The corresponding OLS estimator of  $\beta$  satisfies

$$\sqrt{n}(\hat{\beta}(\mathbf{X}_n) - \beta) \xrightarrow{d} N(0, \sigma_\beta^2) \quad (2.3)$$

with  $\sigma_\beta^2 = 1 - \beta^2$  (c.f. Hamilton, 1994, p. 215). Setting  $\hat{\sigma}_\beta^2(\mathbf{X}_n) = 1 - \hat{\beta}(\mathbf{X}_n)^2$ , an interval for  $\beta x_n$  is typically constructed the following way:

$$\hat{\beta}(\mathbf{X}_n)x_n \pm \Phi^{-1}(\gamma/2) x_n \hat{\sigma}_\beta(\mathbf{X}_n)/\sqrt{n}, \quad (2.4)$$

where  $\Phi^{-1}$  denotes the standard normal quantile function. However, the interval in (2.4) is ill-defined as the terminal observation is treated simultaneously as fixed and random. In essence, researchers typically approximate the law  $\mathcal{L}(\sqrt{n}(\hat{\beta}(\mathbf{X}_n) - \beta)x_n)$  instead of  $\mathcal{L}(\sqrt{n}(\hat{\beta}(\mathbf{X}_n) - \beta)X_n|X_n = x_n)$ . The approximation of the latter appears

rather cumbersome because even the rather simple condition  $X_n = x_n$  has an influence on the whole series  $\mathbf{X}_n$  (Kreiss, 2016). Despite the challenge, Phillips (1979) obtains such approximation based on Edgeworth expansions in the case of  $\varepsilon_t \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2)$ .

A more precarious example, due to its large popularity, is the conditional variance in a generalized autoregressive conditional heteroskedasticity (GARCH) model (Engle, 1982; Bollerslev, 1986). Whereas in the previous AR(1) case it suffices to condition on the terminal observation, the subsequent Example 2 is more extreme as the entire sample contains information about the object of interest.

**Example 2.** Suppose  $\{X_t\}$  follows a GARCH(1, 1) process given by  $X_t = \sigma_t \varepsilon_t$  with

$$\sigma_t^2 = \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2, \quad (2.5)$$

where  $\omega > 0$ ,  $\alpha \geq 0$ ,  $1 > \beta \geq 0$  and  $\{\varepsilon_t\}$  are *i.i.d.* with  $\mathbb{E}[\varepsilon_t] = 0$  and  $\mathbb{E}[\varepsilon_t^2] = 1$ . For simplicity we assume the initial values to be zero:  $X_0^2 = \sigma_0^2 = 0$ . Here, the aim is to construct a  $100(1 - \gamma)\%$  confidence interval for  $\sigma_{n+1|n}^2 = \mathbb{E}[X_{n+1}^2 | \mathbf{X}_n = \mathbf{x}_n]$ , the variance of  $X_{n+1}$  conditional on  $\mathbf{X}_n = \mathbf{x}_n$ . The model's recursive structure implies

$$\sigma_{n+1}^2 = \omega \frac{1 - \beta^n}{1 - \beta} + \alpha \sum_{k=0}^{n-1} \beta^k X_{n-k}^2 + \beta^n (\alpha X_0^2 + \beta \sigma_0^2) \quad (2.6)$$

$$= \omega \frac{1 - \beta^n}{1 - \beta} + \alpha \sum_{k=0}^{n-1} \beta^k X_{n-k}^2, \quad (2.7)$$

which is random. Given the information  $\mathbf{X}_n = \mathbf{x}_n$ , it is non-random and equal to

$$\sigma_{n+1|n}^2 = \psi_n(\mathbf{x}_n; \theta) = \omega \frac{1 - \beta^n}{1 - \beta} + \alpha \sum_{k=0}^{n-1} \beta^k x_{n-k}^2, \quad (2.8)$$

where  $\theta = (\omega, \alpha, \beta)'$ . A natural estimate for the conditional variance is given by

$$\psi_n(\mathbf{x}_n; \hat{\theta}(\mathbf{x}_n)) = \hat{\omega}(\mathbf{x}_n) \frac{1 - \hat{\beta}(\mathbf{x}_n)^n}{1 - \hat{\beta}(\mathbf{x}_n)} + \hat{\alpha}(\mathbf{x}_n) \sum_{k=0}^{n-1} \hat{\beta}(\mathbf{x}_n)^k x_{n-k}^2, \quad (2.9)$$

where  $\hat{\theta}(\mathbf{x}_n) = (\hat{\omega}(\mathbf{x}_n), \hat{\alpha}(\mathbf{x}_n), \hat{\beta}(\mathbf{x}_n))'$  is some estimate for  $\theta$  depending on  $\mathbf{x}_n$ . Typically, an interval is based on the (approximate) quantiles of the ill-defined estimator

$$\hat{\sigma}_{n+1|n}^2 = \psi_n(\mathbf{x}_n; \hat{\theta}(\mathbf{X}_n)) = \hat{\omega}(\mathbf{X}_n) \frac{1 - \hat{\beta}(\mathbf{X}_n)^n}{1 - \hat{\beta}(\mathbf{X}_n)} + \hat{\alpha}(\mathbf{X}_n) \sum_{k=0}^{n-1} \hat{\beta}(\mathbf{X}_n)^k x_{n-k}^2. \quad (2.10)$$

The resulting confidence interval  $I_\gamma^{ILL}(\mathbf{X}_n, \mathbf{x}_n)$  should satisfy exactly or approximately

$$\mathbb{P} \left[ I_\gamma^{ILL}(\mathbf{X}_n, \mathbf{x}_n) \ni \sigma_{n+1|n}^2 \right] \underset{(\approx)}{=} 1 - \gamma. \quad (2.11)$$

However, the statement in (2.11) is not meaningful as the sample is regarded as random and non-random at the same time. A probabilistic statement of the form

$$\mathbb{P} \left[ I_\gamma^{ILL}(\mathbf{X}_n, \mathbf{X}_n) \ni \sigma_{n+1}^2 \middle| \mathbf{X}_n = \mathbf{x}_n \right] \underset{(\approx)}{=} 1 - \gamma \quad (2.12)$$

is likewise nonsensical as conditioning on  $\mathbf{X}_n = \mathbf{x}_n$  means that not only  $\sigma_{n+1}^2 = \sigma_{n+1|n}^2$  is fixed, but also that the confidence interval is equal to the non-random quantity  $I_\gamma^{ILL}(\mathbf{x}_n, \mathbf{x}_n)$  such that the probability that the interval contains the object of interest is either 0 or 1 opposed to (approximately)  $1 - \gamma$ .



### 3 General Setup

#### 3.1 The General Markov Chain Model

Having illustrated the fundamental issue in Section 2, we generalize the structure of the previous examples to a class of Markov chains. Let  $\{X_t\}$  be a real-valued stochastic process defined on the probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ .  $\theta$  denotes a generic parameter vector of length  $r \in \mathbb{N}$  and  $\theta_0$  the true value, unknown to the researcher. Let  $\Theta \subseteq \mathbb{R}^r$  be the corresponding parameter space. Following Boussama et al. (2011), we consider a Markov chain of the form

$$S_t = \varphi(S_{t-1}, X_t; \theta_0), \quad t = 1, 2, \dots \quad (3.1)$$

with some initial value  $S_0$  defined on  $(\Omega, \mathcal{F}, \mathbb{P})$ , where  $\varphi$  is some map  $\varphi : \mathbb{R}^a \times \mathbb{R} \times \Theta \rightarrow \mathbb{R}^a$ . Whereas  $X_t$  is observable by the researcher at time  $t$ ,  $S_t$  may be unobservable or only partially observable. Many stochastic processes are in fact Markov processes. Possibly the simplest case is the AR(1) process in Example 1, for which  $S_t$  equals to  $X_t$ . Subsequently, we revisit Example 2 to establish the mapping for the GARCH(1,1).

**Example 2.** (*continued*) The recursion in (2.5) can be alternatively expressed by

$$\underbrace{\begin{pmatrix} \sigma_t^2 \\ X_t^2 \end{pmatrix}}_{S_t} = \underbrace{\begin{pmatrix} \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2 \\ X_t^2 \end{pmatrix}}_{\varphi(S_{t-1}, X_t; \theta)}. \quad (3.2)$$

The parameter set satisfies  $\Theta \subset (0, \infty) \times [0, \infty) \times [0, 1)$ .

It is a basic exercise to show that the general GARCH( $p, q$ ) with  $p, q \in \mathbb{N}_0$  and autoregressive moving-average (ARMA) models are equally embedded in (3.1). Several

GARCH extensions such as Zakoïan's (1994) threshold GARCH are likewise nested in the Markovian framework. For further details we refer to Beutner et al. (2017).

**Example 3.** Blasques et al. (2016) consider a set of *observation driven models* including Engle and Russell's (1998) autoregressive conditional duration model and score driven models of Harvey (2013) and Creal et al. (2013). The time series is generated by  $X_t \sim p_x(\cdot|f_t; \theta)$  for  $t \in \mathbb{N}$ , where the density  $p_x(\cdot)$  depends on a *time-varying parameter*  $f_t$  and a static parameter vector  $\theta$ .  $f_t$  is updated through a *stochastic recurrence equation* of the form  $f_{t+1} = \phi(X_t, f_t; \theta)$  with initialization  $f_1$  and recurrence function  $\phi(\cdot)$ . The model can be mapped into (3.1) as follows:

$$\underbrace{\begin{pmatrix} f_{t+1} \\ X_t \end{pmatrix}}_{S_t} = \underbrace{\begin{pmatrix} \phi(X_t, f_t; \theta) \\ X_t \end{pmatrix}}_{\varphi(S_{t-1}, X_t; \theta)}. \quad (3.3)$$

Having verified the Markov process structure for special cases, we now turn to the object for which we are interested in constructing a confidence interval with pre-specified conditional coverage probability. The extension to prediction intervals is discussed in Section 5. For a function  $\pi : \mathbb{R}^a \times \Theta \rightarrow \mathbb{R}$  we set  $\lambda_{n+1} = \pi(S_n; \theta_0)$ . Replacing  $t$  by  $n+1$ , equation (2.5) is of this form. Subsequently, we define recursive functions to express  $\lambda_{n+1}$  as a function of the (i) *initial condition*, the (ii) *sample* and the (iii) *true parameter vector*. For  $t \in \mathbb{N}$ , define  $\varphi_t$  recursively by  $\varphi_1 = \varphi$  and  $\varphi_{t+1} : \mathbb{R}^a \times \mathbb{R}^{t+1} \times \Theta \rightarrow \mathbb{R}^a$  with  $\varphi_{t+1}(s, \mathbf{z}_{t+1}; \theta) = \varphi(\varphi_t(s, \mathbf{z}_t; \theta), z; \theta)$  and  $\mathbf{z}_{t+1} = (\mathbf{z}'_t, z)'$ , where  $\mathbf{z}_t \in \mathbb{R}^t$  and  $z \in \mathbb{R}$ . Further, define for  $t \in \mathbb{N}$  the function  $\psi_t : \mathbb{R}^a \times \mathbb{R}^t \times \Theta \rightarrow \mathbb{R}$  by  $\psi_t(s, \mathbf{z}_t; \theta) = \pi(\varphi_t(s, \mathbf{z}_t; \theta); \theta)$ . Then,  $\lambda_{n+1}$  (random) can be expressed by

$$\lambda_{n+1} = \psi_n(S_0, \mathbf{X}_n; \theta_0) \quad (3.4)$$

generalizing (2.6). Conditioning explicitly on  $\mathbf{X}_n = \mathbf{x}_n$  and  $S_0 = s_0$ , it is equal to

$$\lambda_{n+1|n} = \psi_n(s_0, \mathbf{x}_n; \theta_0) , \quad (3.5)$$

which is non-random. Whereas in Example 2 the initial condition was assumed to be known for simplicity, it is usually unknown. In the spirit of Francq and Zakoïan (2004), consider an arbitrary (non-random) initial value denoted by  $s_0^\circ$ . In addition, we denote the estimator of  $\theta_0$  by  $\hat{\theta}(\mathbf{X}_n)$ , a measurable function of the full sample  $\mathbf{X}_n$ , and denote its corresponding estimate by  $\hat{\theta}(\mathbf{x}_n)$ . A natural estimate of  $\lambda_{n+1|n}$  is given by  $\psi_n(s_0^\circ, \mathbf{x}_n; \hat{\theta}(\mathbf{x}_n))$ , whose exemplary counterpart is given in equation (2.9). Let

$$\hat{\lambda}_{n+1}^{ILL} = \psi_n(s_0^\circ, \mathbf{X}_n; \hat{\theta}(\mathbf{X}_n)) \quad (3.6)$$

and observe that it is fixed when conditioning on  $\mathbf{X}_n = \mathbf{x}_n$ . To construct an interval, researchers typically rely on the (approximate) quantiles of the ill-posed estimator

$$\hat{\lambda}_{n+1|n}^{ILL} = \psi_n(s_0^\circ, \mathbf{x}_n; \hat{\theta}(\mathbf{X}_n)) \quad (3.7)$$

resulting in  $I_\gamma^{ILL}(\mathbf{X}_n, \mathbf{x}_n)$ . Along the lines of Example 2, such confidence interval leads to nonsensical probabilistic statements such as

$$\mathbb{P} \left[ I_\gamma^{ILL}(\mathbf{X}_n, \mathbf{x}_n) \ni \lambda_{n+1|n} \right] \underset{(\approx)}{=} 1 - \gamma \quad (3.8)$$

$$\mathbb{P} \left[ I_\gamma^{ILL}(\mathbf{X}_n, \mathbf{X}_n) \ni \lambda_{n+1} \middle| \mathbf{X}_n = \mathbf{x}_n, S_0 = s_0 \right] \underset{(\approx)}{=} 1 - \gamma . \quad (3.9)$$

It is evident that the problem in (3.8) and (3.9) arises due to conditioning. A potential solution would be to treat  $\lambda_{n+1|n}$  as a random quantity in  $\mathbf{X}_n$  and  $S_0$ , e.g.  $\lambda_{n+1}$ . Considering the unconditional coverage probability, any interval would be a *proper*

prediction interval: e.g.  $I_\gamma$  satisfies exactly or approximately  $\mathbb{P}[I_\gamma(\mathbf{X}_n, \mathbf{X}_n) \ni \lambda_{n+1}] = 1 - \gamma$ . While this approach is theoretically sound, such interval relates to what might have happened opposed to what actually happened during the sample period and thus does not possess the *relevance property* of Kabaila (1999) formulated in the context of prediction intervals. Consequently, the approach will lead to considerably wider intervals and as such does not seem to be promising. Instead, researchers typically rely on an argument of two independent processes considered next.

### 3.2 Argument of Two Independent Processes

Although the confidence intervals in (3.8) and (3.9) lead to nonsensical probabilistic statements, researchers frequently argue in favor for constructing intervals in that way. The underlying argument can at least be traced back to Akaike (1969), who studies the prediction of AR time series. It reoccurs in Lewis and Reinsel (1985, p. 394), who assume to have two independent processes with the same stochastic structure: “...the series used for estimation of parameters and the series used for prediction are generated from two independent processes which have the same stochastic structure.” The same argument also appears in Lütkepohl (2005, p. 95) and in Dufour and Taamouti (2010). Let  $\{Y_t\}$  be a process independent of  $\{X_t\}$  and  $S_0$  defined on the same probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  with  $\{Y_t\}$  having the same stochastic structure as  $\{X_t\}$ . In addition to the sample  $\mathbf{X}_n$  of the process  $\{X_t\}$ , suppose there is a sample  $\mathbf{Y}_n = (Y_1, \dots, Y_n)'$  of the process  $\{Y_t\}$ . Then, one can set up an estimator given by

$$\hat{\lambda}_{n+1}^{2IP} = \psi_n(s_0^\circ, \mathbf{X}_n; \hat{\theta}(\mathbf{Y}_n)) \quad (3.10)$$

with randomness induced by  $\hat{\theta}(\mathbf{Y}_n)$  even when conditioning on  $\mathbf{X}_n = \mathbf{x}_n$  such that

$$\hat{\lambda}_{n+1|n}^{2IP} = \psi_n(s_0^\circ, \mathbf{x}_n; \hat{\theta}(\mathbf{Y}_n)) \quad (3.11)$$

is well defined. Throughout this paper, we call (3.10) and its conditional version in (3.11), the 2IP (two independent processes) estimator. Then, an interval can be based on the (approximate) quantiles of  $\hat{\lambda}_{n+1|n}^{2IP}$  resulting in  $I_\gamma^{2IP}(\mathbf{Y}_n, \mathbf{x}_n)$ . It satisfies

$$\mathbb{P}\left[I_\gamma^{2IP}(\mathbf{Y}_n, \mathbf{x}_n) \ni \lambda_{n+1|n}\right] \stackrel{(\approx)}{=} 1 - \gamma \quad (3.12)$$

$$\mathbb{P}\left[I_\gamma^{2IP}(\mathbf{Y}_n, \mathbf{X}_n) \ni \lambda_{n+1} \middle| \mathbf{X}_n = \mathbf{x}_n, S_0 = s_0\right] \stackrel{(\approx)}{=} 1 - \gamma \quad (3.13)$$

or approximately in the case of approximate quantiles. Although this approach is statistically sound, it assumes two independent processes with the same stochastic structure. Phillips (1979) points out that the assumption “is quite unrealistic in practical situations” (p. 241). Indeed, it is difficult to imagine this assumption to be satisfied in any real-life application beyond experimental settings. Moreover, as only one sample realization is available, to compute the estimate of the interval  $I_\gamma^{2IP}(\mathbf{Y}_n, \mathbf{x}_n)$  it is frequently suggested to take  $\mathbf{Y}_n = \mathbf{x}_n$ . This strongly violates the independence assumption. Thus, the 2IP approach appears to be a rather questionable justification for the usual intervals. Nevertheless, it is commonly used in practice and as such, providing an alternative, realistic, justification of intervals that (asymptotically) coincide with these, is of great interest both theoretically and for practitioners.

### 3.3 Sample-split Estimation

An intuitive motivation for the sample-split approach is the successive decline of the influence of past states present in a substantial class of Markov models. This property

permits to split our sample into two (asymptotically) independent subsamples. For  $i = 1, 2$  define the function  $n_i : \mathbb{N} \rightarrow \mathbb{N}$  satisfying  $n_i(n) \rightarrow \infty$  as  $n \rightarrow \infty$  and  $n_1(n) + n_2(n) \leq n$  for all  $n \in \mathbb{N}$ . To simplify notation we abbreviate  $n_i(n)$  by  $n_i$  and set  $\mathbf{X}_n^1 = (X_1, \dots, X_{n_1})'$  and  $\mathbf{X}_n^2 = (X_{n-n_2+1}, \dots, X_n)'$ . The idea is to take the first subsample for the estimation of  $\theta_0$ , whereas the second subsample, consisting of more recent observations, is used for conditioning. The recursive structure of the Markov chain admits to write  $\lambda_{n+1} = \psi_{n_2}(S_{n-n_2}, \mathbf{X}_n^2; \theta_0)$ . Conditioning on  $\mathbf{X}_n^2 = \mathbf{x}_n^2$  and  $S_{n-n_2} = s_{n-n_2}$ , it reduces to  $\lambda_{n+1|n} = \psi_{n_2}(s_{n-n_2}, \mathbf{x}_n^2; \theta_0)$ . Taking a (fixed) arbitrary initial value  $s_{n-n_2}^\circ$ , an alternative estimator is equal to

$$\hat{\lambda}_{n+1}^{SPL} = \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \hat{\theta}(\mathbf{X}_n^1)), \quad (3.14)$$

where  $\hat{\theta}(\mathbf{X}_n^1)$  is an estimator of  $\theta_0$  using the subsample  $\mathbf{X}_n^1$  only. Note that randomness is induced by the estimator  $\hat{\theta}(\mathbf{X}_n^1)$  even when conditioning on the information  $\mathbf{X}_n^2 = \mathbf{x}_n^2$  and  $S_{n-n_2} = s_{n-n_2}$ . Throughout the paper, we call (3.14) the SPL estimator (due to SPLitting). An interval  $I_\gamma^{SPL}(\mathbf{X}_n^1, \mathbf{X}_n^2)$  can be constructed around  $\hat{\lambda}_{n+1}^{SPL}$  such that

$$\mathbb{P}\left[I_\gamma^{SPL}(\mathbf{X}_n^1, \mathbf{X}_n^2) \ni \lambda_{n+1} \mid \mathbf{X}_n^2 = \mathbf{x}_n^2, S_{n-n_2} = s_{n-n_2}\right] \underset{(\approx)}{=} 1 - \gamma. \quad (3.15)$$

This statement does make sense as there is still randomness in  $\hat{\theta}(\mathbf{X}_n^1)$  since  $\mathbf{X}_n^1$  is not conditioned upon, yet the last  $n_2$  values of  $\{X_t\}_{t=1}^n$  and  $S_{n-n_2}$  are fixed such that their randomness is not taken into account.

*Remark 1.* In Section 4 we will discuss how  $n_1$  and  $n_2$  should be chosen from an asymptotic perspective to ensure that our regularity conditions are fulfilled. As we only consider sample splitting as a theoretical approach to validate commonly constructed conditional confidence intervals, these asymptotic guidelines are sufficient for

our purposes and we do not have to consider how to choose  $n_1$  and  $n_2$  in practice. Of course, one could use the sample-split approach in practice to construct confidence intervals. While one would gain (near) independence between the two subsamples, this would come at a cost of estimation precision as fewer observations are used for parameter estimation. For the Gaussian AR(1) setting, Phillips (1979) derives asymptotic expansions for the case where, in our notation,  $n_2 = 1$  and  $n_1 = n - l$  for some  $l \geq 0$ , showing that even in this simple case there is indeed a trade-off as described above and the optimal choice of  $l$  is unclear. An interesting extension of our analysis would therefore be to investigate the optimal choices of  $n_1$  and  $n_2$  to achieve the most accurate confidence intervals in small samples. However, this choice is likely to be highly dependent on the specific model and as such would have to be investigated on a case-by-case basis. This is therefore outside the scope of the current paper.

## 4 Asymptotic Justification of Conditional Confidence Intervals

In this section, we connect the sample-split procedure of Section 3.3 with the two-independent-samples approach of Section 3.2. First, we demonstrate that the notion of weak convergence is inadequate to study asymptotic closeness in this context and discuss the concept of merging. Then, we link the 2IP and the SPL estimator by proving that their conditional distributions merge in probability (Theorem 1). Thereafter, we construct asymptotically valid intervals (Theorem 2) and show that the sample-split intervals coincide asymptotically with the intervals commonly constructed by practitioners (Theorem 3). Last, we state intervals of reduced form and simplified theoretical results under asymptotic normality of the parameter estimator.

To illustrate the shortcomings of weak convergence we revisit Example 1, which shows that studying asymptotic closeness between conditional distributions is often complicated by the absence of a limiting distribution.

**Example 1.** (*continued*) Equation (2.3) entails that  $\sqrt{n}(\hat{\beta}(\mathbf{X}_n) - \beta)x$  converges weakly to  $N(0, \sigma_\beta^2 x^2)$  for an arbitrary *out-of-sample* value  $x \neq 0$ . Further, the result may suggest that  $\mathcal{L}(\sqrt{n}(\hat{\beta}(\mathbf{X}_n) - \beta)X_n | X_n = x_n)$  is asymptotically close to  $N(0, \sigma_\beta^2 x_n^2)$ , which varies with  $n$  through the terminal *in-sample* realization  $x_n$ . Since weak convergence requires a limiting distribution, it is not applicable in this context.

Next, we discuss what closeness means in the absence of a limiting distribution.

## 4.1 Merging and Metrization

D'Aristotile et al. (1988) consider two sequences of probability measures and study cases in which these sequences merge but not necessarily converge. They define merging with respect to the Prokorov metric whereas Dudley (2002, Thm. 11.7.1) shows its metrization by the bounded Lipschitz distance in separable metric spaces. Davydov and Rotar (2009) complement the theory by obtaining equivalence to merging with respect to bounded uniformly continuous functions in complete separable metric spaces. For any real-valued function  $f$  on  $\mathbb{R}^b$ ,  $b \in \mathbb{N}$ , define  $\|f\|_{BL} = \sup_x |f(x)| + \sup_{x \neq y} \frac{|f(x) - f(y)|}{\|x - y\|}$ , where  $\|\cdot\|$  denotes the Euclidean norm, i.e.  $\|A\| = \sqrt{\text{tr}(A'A)}$  for any vector or matrix  $A$ . Further, let  $\mathcal{P}$  be the set of probability measures on  $(\mathbb{R}, \mathcal{B})$ , where  $\mathcal{B}$  denotes the Borel  $\sigma$ -algebra of  $\mathbb{R}$ . For any  $P, Q \in \mathcal{P}$  the bounded Lipschitz metric  $d_{BL} : \mathcal{P} \times \mathcal{P} \rightarrow [0, 1]$  is given by

$$d_{BL}(P, Q) = \sup \left\{ \left| \int f d(P - Q) \right| : \|f\|_{BL} \leq 1 \right\}. \quad (4.1)$$



**Definition 1.** (*Merging*) Two sequences of probability measures  $\{P_n\}, \{Q_n\} \subset \mathcal{P}$  are said to *merge* if and only if  $d_{BL}(P_n, Q_n) \rightarrow 0$  as  $n \rightarrow \infty$ .

It is evident that the definition of merging generalizes the notion of weak convergence since the latter requires a limiting distribution, e.g.  $Q_n = Q$  for all  $n \in \mathbb{N}$ .

## 4.2 Merging of 2IP and SPL in Probability

The goal is to find conditions such that the conditional laws of the 2IP and SPL estimator merge. Let  $m_n$  be a sequence of normalizing constants with  $m_n \rightarrow \infty$  (e.g.  $m_n = \sqrt{n}$ ) and specify  $m_{n_1}$  by replacing  $n$  by  $n_1$ . For  $l, u \in \mathbb{N}$  satisfying  $l \leq u$ , we define the sub  $\sigma$ -algebras  $\mathcal{F}_l^u = \sigma(X_t : l \leq t \leq u)$  and  $\mathcal{G}_l^u = \sigma(S_{l-1}, X_t : l \leq t \leq u)$ , where  $\mathcal{G}_l^u$  includes the respective initial condition. We denote the *regular conditional distributions* of the 2IP and SPL estimator by

$$P_n^{2IP}(\cdot | \mathcal{G}_1^n) = P[m_n(\hat{\lambda}_{n+1}^{2IP} - \lambda_{n+1}) \in \cdot | \mathcal{G}_1^n] \quad (4.2)$$

$$P_n^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n) = P[m_n(\hat{\lambda}_{n+1}^{SPL} - \lambda_{n+1}) \in \cdot | \mathcal{G}_{n-n_2+1}^n], \quad (4.3)$$

respectively.<sup>2</sup> We refer to Dudley (2002, p. 341) for the definition of the regular conditional distribution such that  $P_n^{2IP}(\cdot | \mathcal{G}_1^n)(\omega) \in \mathcal{P}$  and  $P_n^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n)(\omega) \in \mathcal{P}$  for every  $\omega \in \Omega$ . For their existence see Dudley (2002, Thm. 10.2.2). Further, since  $\omega \rightarrow P_n^{2IP}(\cdot | \mathcal{G}_1^n)(\omega)$  and  $\omega \rightarrow P_n^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n)(\omega)$  can be viewed as random variables it follows that  $d_{BL}(P_n^{2IP}(\cdot | \mathcal{G}_1^n), P_n^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n))$  is a random variable (cf. Klenke, 2014, Lem. 6.1). Henceforth, we consider merging in probability defined as follows.

**Definition 2.** (*Merging in Probability*) Two sequences of regular conditional distributions  $P_n$  and  $Q_n$  defined on  $\mathcal{B} \times \Omega$  are said to *merge in probability* if and only if

---

<sup>2</sup>Note that it does not suffice to condition on  $\mathcal{F}_1^n$  (or  $\mathcal{F}_{n-n_2+1}^n$  respectively) here as the object of interest would not be fixed, but random due to variation in the initial condition.

$d_{BL}(P_n, Q_n) \xrightarrow{p} 0$  as  $n \rightarrow \infty$ , where “ $\xrightarrow{p}$ ” denotes “convergence in probability”.

*Remark 2.* There is another, yet not equivalent, way to define merging in probability as introduced in Belyaev and Sjöstedt-De Luna (2000). Some of our intermediate lemmas on merging in probability are closely related to their results, see the proofs in Appendix A for details.

To write the subsequent assumptions in compact form we employ the usual stochastic order symbols  $O_p$  and  $o_p$ . We assume  $\theta_0$  belongs to an open set  $\Theta^\star \subset \mathring{\Theta}$ , where  $\mathring{\Theta}$  stands for the interior of  $\Theta$ , and denote the set of all bounded, real-valued Lipschitz functions on  $\mathbb{R}^r$  by  $BL = \{h : \mathbb{R}^r \rightarrow \mathbb{R} : \|h\|_{BL} < \infty\}$ . We make the following assumptions regarding the 2IP estimator.

**Assumption 1.** (*2IP Estimator*)

1.1 (*Estimator*)  $m_n(\hat{\theta}(\mathbf{X}_n) - \theta_0) \xrightarrow{d} P_{\theta_0, \xi_0}$  as  $n \rightarrow \infty$  for some law  $P_{\theta_0, \xi_0}$  belonging to a parametric family  $\{P_{\theta, \eta} | \theta \in \Theta, \eta \in \Xi\}$ ;

1.2 (*Existence & Independence*)

(a)  $\{Y_t\}$  is a process defined on  $(\Omega, \mathcal{F}, \mathbb{P})$  with  $\mathcal{L}(\{Y_t\}) = \mathcal{L}(\{X_t\})$ ;

(b)  $\{Y_t\}$  is independent of  $\{X_t\}$  and  $S_0$ ;

1.3 (*Differentiability*)  $\psi_n(\cdot, \cdot; \theta)$  is continuous on  $\Theta$  and twice differentiable on  $\mathring{\Theta}$ ;

1.4 (*Initial Cond.*)  $m_n(\psi_n(S_0, \mathbf{X}_n; \theta_0) - \psi_n(s_0^\circ, \mathbf{X}_n; \theta_0)) = o_p(1)$  for all  $s_0^\circ \in \mathbb{R}^a$ ;

1.5 (*Hessian*)  $\sup_{\theta \in \Theta^\star} \left\| \frac{\partial^2 \psi_n(s_0^\circ, \mathbf{X}_n; \theta)}{\partial \theta \partial \theta'} \right\| = O_p(1)$  for all  $s_0^\circ \in \mathbb{R}^a$ ;

1.6 (*Gradient*)  $\left\| \frac{\partial \psi_n(s_0^\circ, \mathbf{X}_n; \theta_0)}{\partial \theta} \right\| = O_p(1)$  for all  $s_0^\circ \in \mathbb{R}^a$ .

The parameter estimator is assumed to converge weakly in 1.1. The limiting distribution may not only depend on  $\theta_0$ , but potentially also on some nuisance parameter vector  $\xi_0$ . Assumption 1.2 states the unrealistic two-independent-processes assumption, that needs to be relaxed. The differentiability assumption in 1.3 admits a stochastic Taylor expansion of the form

$$m_n(\hat{\lambda}_{n+1}^{2IP} - \lambda_{n+1}) = \frac{\partial \psi_n(s_0^\circ, \mathbf{X}_n; \theta_0)}{\partial \theta'} m_n(\hat{\theta}(\mathbf{Y}_n) - \theta_0) + R_n^{2IP}, \quad (4.4)$$

where  $R_n^{2IP}$  is a remainder term. Assumption 1.4 ensures the initial condition to be asymptotically negligible and is needed together with Assumption 1.5 to show that  $R_n^{2IP}$  is of order  $o_p(1)$ . Assumption 1.6 states that the weights of the rescaled parameter estimator in (4.4) are bounded in probability. Regarding the SPL estimator we state a similar set of assumptions, but replace the unrealistic assumption of two independent processes by a weak dependence condition, which allows to split our sample into two (asymptotically) independent subsamples.

**Assumption 2.** (*SPL Estimator*)

2.1 (*Estimator*)  $Z_n^{SPL} = m_n(\hat{\theta}(\mathbf{X}_n^1) - \theta_0) \xrightarrow{d} P_{\theta_0, \xi_0}$  as  $n \rightarrow \infty$ ;

2.2 (*Weak Dependence*)  $\{X_t\}$  satisfies for each  $h \in BL$

$$\int h d\left(P_{Z_n}^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n) - P_{Z_n}^{SPL}\right) \xrightarrow{p} 0 \quad \text{as} \quad n \rightarrow \infty,$$

where  $n_3 = n - n_1 - n_2$ ,  $P_{Z_n}^{SPL}$  denotes the unconditional distribution of  $Z_n^{SPL}$  and  $P_{Z_n}^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n)$  the corresponding conditional law given  $\mathcal{G}_{n-n_2+1}^n$ ;

2.3 (*Differentiability*)  $\psi_{n_2}(\cdot, \cdot; \theta)$  is continuous on  $\Theta$  and twice differentiable on  $\mathring{\Theta}$ ;

2.4 (*Initial Cond.*)  $m_n(\psi_{n_2}(S_{n-n_2}, \mathbf{X}_n^2; \theta_0) - \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \theta_0)) = o_p(1)$  for all  $s_{n-n_2}^\circ \in \mathbb{R}^a$ ;

2.5 (*Hessian*)  $\sup_{\theta \in \Theta^*} \left\| \frac{\partial^2 \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \theta)}{\partial \theta \partial \theta'} \right\| = O_p(1)$  for all  $s_{n-n_2}^\circ \in \mathbb{R}^a$ ;

2.6 (*Gradient*)  $\left\| \frac{\partial \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \theta_0)}{\partial \theta} \right\| = O_p(1)$  for all  $s_{n-n_2}^\circ \in \mathbb{R}^a$ .

In Assumption 2.1 it seems natural to scale by  $m_{n_1}$ . Through scaling by  $m_n$  instead, we implicitly assume that  $\frac{m_{n_1}}{m_n} \rightarrow 1$ , which typically reduces to  $\frac{n_1}{n} \rightarrow 1$ . The weak dependence condition in 2.2 is met by numerous Markov processes. Intuitively,  $(X_1, \dots, X_{n_1})$  and  $(S'_{n-n_2}, X_{n-n_2+1}, \dots, X_n)$  approach independence as their temporal distance  $n_3$  increases. We illustrate one particular case in the following remark.

*Remark 3.* Suppose  $(S'_t, X_t)$  is strong mixing, that is  $\alpha(k) = \sup_{j \in \mathbb{Z}} \alpha(\mathcal{A}_{-\infty}^j, \mathcal{A}_{j+k}^\infty) \rightarrow 0$  as  $k \rightarrow \infty$  with  $\alpha(\mathcal{C}, \mathcal{D}) = \sup_{C \in \mathcal{C}, D \in \mathcal{D}} |\mathbb{P}(C \cap D) - \mathbb{P}(C)\mathbb{P}(D)|$ , where  $\mathcal{A}_l^u = \sigma((S'_t, X_t), l \leq t \leq u)$  with  $l, u \in \bar{\mathbb{Z}}$  (Doukhan, 1994). For  $h \in BL$  with  $\|h\|_{BL} \leq K$  and for all  $\epsilon > 0$ , Markov's and Ibragimov's inequality (cf. Hall and Heyde, 1980, Thm. A.5) imply

$$\begin{aligned} & \mathbb{P} \left[ \left| \int h d(P_{Z_n}^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n) - P_{Z_n}^{SPL}) \right| \geq \epsilon \right] \\ & \leq \frac{1}{\epsilon} \mathbb{E} \left| \int h d(P_{Z_n}^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n) - P_{Z_n}^{SPL}) \right| \\ & = \frac{1}{\epsilon} \mathbb{Cov} \left[ h(Z_n^{SPL}), \text{sign} \left\{ \int h d(P_{Z_n}^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n) - P_{Z_n}^{SPL}) \right\} \right] \\ & \leq \frac{4K}{\epsilon} \alpha(\mathcal{A}_1^{n_1}, \mathcal{A}_{n-n_2+1}^n) \leq \frac{4K}{\epsilon} \alpha(n_3). \end{aligned}$$

It follows that Assumption 2.2 is met as  $n_3 \rightarrow \infty$ .

Further, Assumption 2.4 implicitly limits the choice of  $n_2$ . For models exhibiting an exponential decay in memory, it typically suffices to take  $\frac{n_2}{\log n} \rightarrow \infty$  (see Beutner et al., 2017, Eq. (2.9)). The other assumptions are analogous to the corre-

sponding assumptions of the 2IP estimator such that we can write

$$m_n(\hat{\lambda}_{n+1}^{SPL} - \lambda_{n+1}) = \frac{\partial \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \theta_0)}{\partial \theta'} m_n(\hat{\theta}(\mathbf{X}_n^1) - \theta_0) + R_n^{SPL}, \quad (4.5)$$

where  $R_n^{SPL}$  is a remainder. To show asymptotic closeness between the conditional laws of the 2IP and the SPL estimator, we need the gradients to merge in probability.

**Assumption 3.** (*Merging Gradients*) We assume for all  $s_0^\circ, s_{n-n_2}^\circ \in \mathbb{R}^a$

$$\left\| \frac{\partial \psi_n(s_0^\circ, \mathbf{X}_n; \theta_0)}{\partial \theta} - \frac{\partial \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \theta_0)}{\partial \theta} \right\| \xrightarrow{p} 0.$$

Assumption 3 ensures that the weights corresponding to the rescaled parameter estimators in (4.4) and (4.5) are of comparable size. For common time series models the assumption turns out to be not restrictive and can be straightforwardly verified. Intuitively, the assumption is closely related to the asymptotic negligibility of the initial condition. From the recursive structure of  $\psi_n(\cdot)$  we can deduce that  $\psi_n(s_0^\circ, \mathbf{X}_n; \theta) = \psi_{n_2}(\tilde{S}_{n-n_2}^\circ, \mathbf{X}_n^2; \theta)$ , for a (random) initial condition  $\tilde{S}_{n-n_2}^\circ$  determined by  $s_0^\circ$  and  $\mathbf{X}_n^1$ . Hence, showing that Assumption 3 holds then amounts to proving the asymptotic negligibility for the gradient of the initial conditions  $s_0^\circ$  and  $\tilde{S}_{n-n_2}^\circ$ .

Assumptions 1 to 3 are met by the AR and GARCH processes considered in Examples 1 and 2. A detailed verification of each assumption under mild conditions is provided in Beutner et al. (2017). Building upon Assumptions 1 to 3, we state the following theorem.

**Theorem 1.** (*Merging of 2IP and SPL*) Under Assumptions 1 to 3,  $P_n^{2IP}(\cdot | \mathcal{G}_1^n)$  and  $P_n^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n)$  merge in probability.

Having established asymptotic closeness between the conditional laws  $P_n^{2IP}(\cdot | \mathcal{G}_1^n)$  and  $P_n^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n)$ , we now turn to the construction of asymptotic intervals.

### 4.3 Interval Construction

Henceforth, for any cumulative distribution function (cdf)  $F$  we write  $F^{-1}$  to denote its generalized inverse given by  $F^{-1}(u) = \inf \{ \tau \in \mathbb{R} : F(\tau) \geq u \}$ . For  $\tau \in \mathbb{R}$ , define

$$F_n^{2IP}(\tau | \mathcal{G}_1^n) = P_n^{2IP}((-\infty, \tau] | \mathcal{G}_1^n) \quad (4.6)$$

$$F_n^{SPL}(\tau | \mathcal{G}_{n-n_2+1}^n) = P_n^{SPL}((-\infty, \tau] | \mathcal{G}_{n-n_2+1}^n), \quad (4.7)$$

the conditional cdfs of  $m_n(\hat{\lambda}_{n+1}^{2IP} - \lambda_{n+1})$  given  $\mathcal{G}_1^n$  and  $m_n(\hat{\lambda}_{n+1}^{SPL} - \lambda_{n+1})$  given  $\mathcal{G}_{n-n_2+1}^n$ . A confidence interval for  $\lambda_{n+1}$  based on quantiles of (4.6) or (4.7) is typically infeasible as the underlying conditional laws are unknown for finite  $n$ . Frequently, one relies on an asymptotic approximation while replacing the unknown parameters by consistent estimators. Let  $\widehat{F}_n^{2IP}(\cdot)$  and  $\widehat{F}_n^{SPL}(\cdot)$  be the estimators of (4.6) and (4.7), respectively, motivated by the plug-in principle and denote their corresponding laws by  $\widehat{P}_n^{2IP}(\cdot)$  and  $\widehat{P}_n^{SPL}(\cdot)$ . The specific form of conditional estimators primarily depend on  $P_{\theta_0, \xi_0}$ , the limiting distribution of the parameter estimator. In Section 4.4, we provide explicit expressions when  $P_{\theta_0, \xi_0}$  is multivariate normal. For the general construction, we refer to relations (A.3) and (A.4) in Appendix A and the explanations preceding these relations. Based on the 2IP approach, we consider an interval of the form

$$I_\gamma^{2IP}(\mathbf{Y}_n, \mathbf{X}_n) = \left[ \hat{\lambda}_{n+1}^{2IP} - \frac{\widehat{F}_n^{2IP}{}^{-1}(1 - \gamma_2)}{m_n}, \hat{\lambda}_{n+1}^{2IP} - \frac{\widehat{F}_n^{2IP}{}^{-1}(\gamma_1)}{m_n} \right], \quad (4.8)$$

where  $\gamma_1, \gamma_2 \in [0, 1)$  satisfy  $\gamma = \gamma_1 + \gamma_2$ . We typically take  $\gamma_1 = \gamma_2 = \gamma/2$  such that the interval is equal-tailed. Similarly, we construct the following sample split interval:

$$I_\gamma^{SPL}(\mathbf{X}_n^1, \mathbf{X}_n^2) = \left[ \hat{\lambda}_{n+1}^{SPL} - \frac{\widehat{F}_n^{SPL}{}^{-1}(1 - \gamma_2)}{m_n}, \hat{\lambda}_{n+1}^{SPL} - \frac{\widehat{F}_n^{SPL}{}^{-1}(\gamma_1)}{m_n} \right]. \quad (4.9)$$

To achieve correct coverage, we need that  $\widehat{P}_n^{2IP}(\cdot)$  and  $P_n^{2IP}(\cdot|\mathcal{G}_1^n)$  merge in probability and likewise for SPL. A sufficient condition for this in our setting is that we can consistently estimate the asymptotic distribution of the parameter estimator,  $P_{\theta_0, \xi_0}$ , by the plug-in estimator. This is formulated in Assumption 4 below.

**Assumption 4.** (*Plug-in Estimator*)

$$4.1 \text{ (2IP)} \quad \int h dP_{\hat{\theta}(\mathbf{X}_n), \hat{\xi}(\mathbf{X}_n)} \xrightarrow{p} \int h dP_{\theta_0, \xi_0} \text{ as } n \rightarrow \infty \text{ for each } h \in BL;$$

$$4.2 \text{ (SPL)} \quad \int h dP_{\hat{\theta}(\mathbf{X}_n^1), \hat{\xi}(\mathbf{X}_n^1)} \xrightarrow{p} \int h dP_{\theta_0, \xi_0} \text{ as } n \rightarrow \infty \text{ for each } h \in BL.$$

The verification of Assumption 4 is a standard step in asymptotic analysis. For instance, given some consistent estimators for  $\theta_0$  and  $\xi_0$ , it follows from the continuous mapping theorem if  $P_{\theta, \xi}$  is continuous in  $\theta$  and  $\xi$ . Equivalently, using a consistent bootstrap procedure also guarantees Assumption 4 and consequently the merging of  $\widehat{P}_n^{2IP}(\cdot)$  and  $P_n^{2IP}(\cdot|\mathcal{G}_1^n)$  in probability (similar for SPL). The following theorem states the intervals' asymptotic validity.

**Theorem 2.** (*Asymptotic Coverage*)

1. (a) Under Assumption 1 and 4,  $P_n^{2IP}(\cdot|\mathcal{G}_1^n)$  and  $\widehat{P}_n^{2IP}(\cdot)$  merge in probability.
- (b) If in addition  $\widehat{F}_n^{2IP}(\cdot)$  is stochastically uniformly equicontinuous, then

$$\mathbb{P}\left[I_\gamma^{2IP}(\mathbf{Y}_n, \mathbf{X}_n) \ni \lambda_{n+1} \middle| \mathcal{G}_1^n\right] \xrightarrow{p} 1 - \gamma. \quad (4.10)$$

2. (a) Under Assumption 2 and 4,  $P_n^{SPL}(\cdot|\mathcal{G}_{n-n_2+1}^n)$  and  $\widehat{P}_n^{SPL}(\cdot)$  merge in probability.
- (b) If in addition  $\widehat{F}_n^{SPL}(\cdot)$  is stochastically uniformly equicontinuous, then

$$\mathbb{P}\left[I_\gamma^{SPL}(\mathbf{X}_n^1, \mathbf{X}_n^2) \ni \lambda_{n+1} \middle| \mathcal{G}_{n-n_2+1}^n\right] \xrightarrow{p} 1 - \gamma. \quad (4.11)$$

*Remark 4.* The dominated convergence theorem (c.f. Dudley, 2002, Thm. 4.3.5) entails that the  $\sigma$ -algebras in (4.10) and (4.11) can be replaced by  $\mathcal{F}_1^n$  and  $\mathcal{F}_{n-n_2+1}^n$ , respectively. Whereas  $\mathcal{G}_1^n$  and  $\mathcal{G}_{n-n_2+1}^n$  still depend on initial conditions, that are typically unobserved, the  $\sigma$ -algebras  $\mathcal{F}_1^n$  and  $\mathcal{F}_{n-n_2+1}^n$  are fully known.

Notice that Assumption 3, which links the conditional laws of the 2IP and the SPL estimator, is not required in Theorem 2 since we consider both estimators separately. Researchers typically motivate their conditional confidence interval by  $I_\gamma^{2IP}$  using (4.10) and compute an estimate of the form  $I_\gamma^{2IP}(\mathbf{x}_n, \mathbf{x}_n)$  as only one sample realization is available. This, of course, strongly violates the independence assumption of  $\{X_t\}$  and  $\{Y_t\}$ . Nevertheless, replacing  $\mathbf{Y}_n$  by  $\mathbf{X}_n$  in equation (4.8), leads to

$$I_\gamma^{ILL}(\mathbf{X}_n, \mathbf{X}_n) = \left[ \hat{\lambda}_{n+1}^{ILL} - \frac{\widehat{F}_n^{ILL}{}^{-1}(1 - \gamma_2)}{m_n}, \hat{\lambda}_{n+1}^{ILL} - \frac{\widehat{F}_n^{ILL}{}^{-1}(\gamma_1)}{m_n} \right], \quad (4.12)$$

where  $\widehat{F}_n^{ILL}(\cdot)$  is the conditional cdf of  $\widehat{P}_n^{ILL}(\cdot)$ , and  $\widehat{P}_n^{ILL}(\cdot)$  is defined in relation (A.5) and the text preceding it. Whereas a confidence interval of the form  $I_\gamma^{ILL}(\mathbf{X}_n, \mathbf{X}_n)$  is nonsensical with regard to (3.8) and (3.9), its corresponding estimate  $I_\gamma^{ILL}(\mathbf{x}_n, \mathbf{x}_n)$  is usually numerically close to  $I_\gamma^{SPL}(\mathbf{x}_n^1, \mathbf{x}_n^2)$ . We establish the asymptotic equivalence between the two intervals in the following theorem.

**Theorem 3.** (*Asymptotic Equivalence Confidence Intervals*)

1. (*Location*) If Assumption 1 and 2 hold, then  $\hat{\lambda}_{n+1}^{ILL} - \hat{\lambda}_{n+1}^{SPL} \xrightarrow{p} 0$ .
2. (*Length*) Under the assumptions of Theorem 1 and 2 and  $\widehat{F}_n^{SPL}{}^{-1}(\cdot)$  being stochastically pointwise continuous at  $u = \gamma_1, 1 - \gamma_2$ , we have

$$\widehat{F}_n^{ILL}{}^{-1}(u) - \widehat{F}_n^{SPL}{}^{-1}(u) \xrightarrow{p} 0. \quad (4.13)$$



The first implication states that the locations of the two intervals coincide asymptotically. The second statement establishes asymptotic closeness of the selected quantiles such that the scaled lengths of the intervals in (4.9) and (4.12) coincide asymptotically. We conclude that our sample-split interval  $I_\gamma^{SPL}(\mathbf{x}_n^1, \mathbf{x}_n^2)$  coincides asymptotically with the standard interval  $I_\gamma^{ILL}(\mathbf{x}_n, \mathbf{x}_n)$ . Thus, we provide a realistic justification for the intervals commonly constructed by practitioners.

#### 4.4 Interval Construction Under Normality

In this subsection we present intervals of reduced form and simplified theoretical results under asymptotic normality of the parameter estimator.

**Assumption 5.** (*Normality*) We assume  $P_{\theta_0, \xi_0} = N(0, \Upsilon_0)$  with  $\Upsilon_0 = \Upsilon(\theta_0, \xi_0)$  and there exist  $\hat{\Upsilon}(\mathbf{X}_n)$  and  $\hat{\Upsilon}(\mathbf{X}_n^1)$  converging in probability to  $\Upsilon_0$ .

Usually, the covariance estimator is obtained by inserting consistent estimators for  $\theta_0$  and  $\xi_0$  into  $\Upsilon_0$ . Following the plug-in principle, we estimate  $P_n^{2IP}(\cdot | \mathcal{G}_1^n)$  by a normal distribution with mean 0 and variance  $\hat{v}_n^{2IP} = \frac{\partial \psi_n(s_0^\circ, \mathbf{X}_n; \hat{\theta}(\mathbf{Y}_n))}{\partial \theta'} \hat{\Upsilon}(\mathbf{Y}_n) \frac{\partial \psi_n(s_0^\circ, \mathbf{X}_n; \hat{\theta}(\mathbf{Y}_n))}{\partial \theta}$  such that  $\widehat{P}_n^{2IP}(\cdot)$  equals to  $N(0, \hat{v}_n^{2IP})$ . Then, the interval in (4.8) simplifies to

$$I_\gamma^{2IP}(\mathbf{Y}_n, \mathbf{X}_n) = \left[ \hat{\lambda}_{n+1}^{2IP} - \frac{\sqrt{\hat{v}_n^{2IP}} \Phi^{-1}(1 - \gamma_2)}{m_n}, \hat{\lambda}_{n+1}^{2IP} - \frac{\sqrt{\hat{v}_n^{2IP}} \Phi^{-1}(\gamma_1)}{m_n} \right]. \quad (4.14)$$

Similarly, for the sample-split approach we consider  $\widehat{P}_n^{SPL}(\cdot)$  equals to  $N(0, \hat{v}_n^{SPL})$  with  $\hat{v}_n^{SPL} = \frac{\partial \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \hat{\theta}(\mathbf{X}_n^1))}{\partial \theta'} \hat{\Upsilon}(\mathbf{X}_n^1) \frac{\partial \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \hat{\theta}(\mathbf{X}_n^1))}{\partial \theta'}$  such that (4.9) reduces to

$$I_\gamma^{SPL}(\mathbf{X}_n^1, \mathbf{X}_n^2) = \left[ \hat{\lambda}_{n+1}^{SPL} - \frac{\sqrt{\hat{v}_n^{SPL}} \Phi^{-1}(1 - \gamma_2)}{m_n}, \hat{\lambda}_{n+1}^{SPL} - \frac{\sqrt{\hat{v}_n^{SPL}} \Phi^{-1}(\gamma_1)}{m_n} \right]. \quad (4.15)$$

In Appendix B.2 we show that if the variance estimator is bounded away from zero in probability, e.g.  $1/\hat{v}_n^{2IP} = O_p(1)$ , the corresponding conditional cdf, e.g.  $\widehat{F}_n^{2IP}(\cdot) =$

$\Phi(\cdot / \sqrt{\hat{v}_n^{2IP}})$ , is stochastically uniform equicontinuous. Therefore, the asymptotic validity of both intervals can be deduced from Theorem 2.

**Corollary 1.** (*Asymptotic Coverage under Normality*)

1. (a) Under Assumption 1 and 5,  $P_n^{2IP}(\cdot | \mathcal{G}_1^n)$  and  $N(0, \hat{v}_n^{2IP})$  merge in probability.

(b) If in addition  $1/\hat{v}_n^{2IP} = O_p(1)$ , then

$$\mathbb{P}\left[I_\gamma^{2IP}(\mathbf{Y}_n, \mathbf{X}_n) \ni \lambda_{n+1} \middle| \mathcal{G}_1^n\right] \xrightarrow{p} 1 - \gamma. \quad (4.16)$$

2. (a) Under Assumption 2 and 5,  $P_n^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n)$  and  $N(0, \hat{v}_n^{SPL})$  merge in probability.

(b) If in addition  $1/\hat{v}_n^{SPL} = O_p(1)$ , then

$$\mathbb{P}\left[I_\gamma^{SPL}(\mathbf{X}_n^1, \mathbf{X}_n^2) \ni \lambda_{n+1} \middle| \mathcal{G}_{n-n_2+1}^n\right] \xrightarrow{p} 1 - \gamma. \quad (4.17)$$

Bounding the variance estimator away from zero in probability to establish that the conditional coverage probability converges to  $1 - \gamma$  in probability has intuitive appeal: as  $\hat{v}_n^{2IP}$  approaches zero,  $N(0, \hat{v}_n^{2IP})$  becomes degenerate while the interval in (4.14) collapses (similar for SPL).

Violating Assumption 1.2(b) by replacing  $\mathbf{Y}_n$  by  $\mathbf{X}_n$  in (4.14) leads to

$$I_\gamma^{ILL}(\mathbf{X}_n, \mathbf{X}_n) = \left[ \hat{\lambda}_{n+1}^{ILL} - \frac{\sqrt{\hat{v}_n^{ILL}} \Phi^{-1}(1 - \gamma_2)}{m_n}, \hat{\lambda}_{n+1}^{ILL} - \frac{\sqrt{\hat{v}_n^{ILL}} \Phi^{-1}(\gamma_1)}{m_n} \right] \quad (4.18)$$

with  $\hat{v}_n^{ILL} = \frac{\partial \psi_n(s_0^\circ, \mathbf{X}_n; \hat{\theta}(\mathbf{X}_n))}{\partial \theta'} \hat{\Upsilon}(\mathbf{X}_n) \frac{\partial \psi_n(s_0^\circ, \mathbf{X}_n; \hat{\theta}(\mathbf{X}_n))}{\partial \theta}$ . In Appendix B.2 we prove that the  $\hat{v}_n^{SPL}$  is bounded in probability, which in turn implies that quantile function  $\widehat{F_n^{SPL}}^{-1}(\cdot) = \sqrt{\hat{v}_n^{SPL}} \Phi^{-1}(\cdot)$  is stochastically pointwise equicontinuous at any  $u \in \mathbb{R}$ .

Hence, Theorem 3 applies. Whereas the first statement of the theorem remains unaffected, its second statement reads as follows under normality.

**Corollary 2.** (*Length under Normality*) Under the assumptions of Theorem 1 and Corollary 1, we have for  $u = \gamma_1, 1 - \gamma_2$

$$\sqrt{\hat{v}_n^{ILL}}\Phi^{-1}(u) - \sqrt{\hat{v}_n^{SPL}}\Phi^{-1}(u) \xrightarrow{p} 0. \quad (4.19)$$

## 5 Prediction Intervals

The preceding sections have focused purely on the construction of conditional confidence intervals to account for parameter uncertainty. Regarding prediction, a second source of uncertainty arises, that corresponds to the model's innovation process. In this setting, parameter estimation is typically disregarded in textbooks as the stochastic fluctuation stemming from the estimation procedure is generally dominated by the stochastic fluctuation of the innovations. Although the resulting prediction intervals may be asymptotically valid, they are typically characterized by under-coverage in finite samples. In response, Pan and Politis (2016a) introduce the concept of *asymptotic pertinence* to evaluate distribution approximations that account for the two sources of randomness, innovation and parameter estimation uncertainty, according to their general orders of magnitude. Whereas Kunitomo and Yamamoto (1985) and Samaranayake and Hasza (1988) study properties of the *unconditional* law of the forecast error, we focus on its *conditional* distribution to conform with the relevance property of Kabaila (1999). The fundamental issue also arises when considering prediction if one attempts to account for parameter uncertainty. To illustrate this point, we revisit the introductory examples and write  $*$  to denote the convolution operator.

**Example 1.** (*continued*) Prediction intervals for the AR are often constructed

around the point estimate for the conditional mean. The conditional distribution of the forecast error decomposes into

$$\mathcal{L}(X_{n+1} - \hat{\beta}(\mathbf{X}_n)X_n | X_n = x_n) = \mathcal{L}(\beta X_n - \hat{\beta}(\mathbf{X}_n)X_n | X_n = x_n) * \mathcal{L}(\varepsilon_{n+1}), \quad (5.1)$$

corresponding to estimation and innovation uncertainty, respectively. As argued above, an approximation of  $\mathcal{L}(\beta X_n - \hat{\beta}(\mathbf{X}_n)X_n | X_n = x_n)$  appears rather cumbersome. In the special case of  $\varepsilon_t \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2)$ , Phillips (1979, Thm. 3) derives an approximation for (5.1) based on Edgeworth expansions.

**Example 2.** (*continued*) Suppose we are interested in providing a prediction interval for  $X_{n+1}^2$  in the GARCH(1,1). Conditioning on  $\mathbf{X}_n = \mathbf{x}_n$ , a natural estimate of  $X_{n+1}^2$  is  $\hat{\sigma}_{n+1|n}^{2\,ILL}$ , since  $\sigma_{n+1|n}^2$  is its expected value given information up to time  $n$ . As

$$\begin{aligned} \mathcal{L}(X_{n+1}^2 - \hat{\sigma}_{n+1}^{2\,ILL} | \mathbf{X}_n = \mathbf{x}_n) &= \mathcal{L}(\sigma_{n+1}^2 - \hat{\sigma}_{n+1}^{2\,ILL} | \mathbf{X}_n = \mathbf{x}_n) \\ &* \mathcal{L}(\sigma_{n+1|n}^2(\varepsilon_{n+1}^2 - 1)), \end{aligned} \quad (5.2)$$

where  $\hat{\sigma}_{n+1}^{2\,ILL}$  is defined analogously to (3.6), the desired prediction interval, say  $J_\gamma^{ILL}$ , leads to a sensible probabilistic statement due to variability in  $\varepsilon_{n+1}^2$ :

$$\mathbb{P}\left[X_{n+1}^2 \in J_\gamma^{ILL}(\mathbf{X}_n, \mathbf{X}_n) \middle| \mathbf{X}_n = \mathbf{x}_n\right] \underset{(\approx)}{=} 1 - \gamma. \quad (5.3)$$

However, it cannot incorporate parameter uncertainty either, since the conditional law  $\mathcal{L}(\sigma_{n+1}^2 - \hat{\sigma}_{n+1}^{2\,ILL} | \mathbf{X}_n = \mathbf{x}_n) = \mathcal{L}(\sigma_{n+1|n}^2 - \hat{\sigma}_{n+1|n}^{2\,ILL})$  is degenerate.

In his textbook Pesaran (2015) resorts to a Bayesian-akin approach to avoid the fundamental issue in forecasting. He argues that  $\theta$ , although “fixed at the estimation stage ... is viewed best as a random variable at the forecasting stage” (p. 389).

Consequently, he assigns some posterior distribution to  $\theta$  motivated by an uninformed prior. Treating  $\theta$  not fixed but random, the fundamental issue does not arise, however combining a frequentist view with a Bayesian method does not seem to be coherent.

Barndorff-Nielsen and Cox (1996) require the existence of a transitive statistic  $U = U(\mathbf{X}_n)$  of fixed low dimension to establish conditional independence between the sample  $\mathbf{X}_n$  and their considered future random variable given  $U = u$ . Vidoni (2004, 2009a, 2009b, 2017), Kabaila (1999), Kabaila and He (2004), and Kabaila and Syuhada (2008, 2010) extend their approach and derive improved prediction intervals. Although these methods absorb an additional  $O(n^{-1})$  term in the associated conditional coverage probability to account for parameter uncertainty, there are several drawbacks associated with them: the innovation distribution needs typically be specified (e.g. Gaussian), the results apply only to a limited set of estimators (e.g. maximum likelihood) and their framework can only incorporate finite autoregressive components (e.g. AR(p)).

Assuming two independent processes with the same stochastic structure, using one for prediction and one for the estimation of the parameters, alleviates the fundamental issue faced in the continued Examples 1 and 2. As the conditional distributions of the 2IP and SPL estimators merge in probability by Theorem 1, the 2IP assumption can be avoided by following a sample-split approach as described in Section 3.3.

## 6 Conclusion

In the paper at hand, we study the construction of confidence intervals for conditional objects such as conditional means or conditional variances. Two examples illustrate the fundamental issue that arises in the process of taking parameter uncertainty into account. It stems from the facts that on one hand one must condition on the sample

as the past informs about the present and future, yet on the other hand one must allow the data up to now to be treated as random to account for estimation uncertainty. Assuming two independent processes with the same stochastic structure, where one is used for conditioning and one for the estimation of the parameters, bypasses this issue, but the assumption itself is violated in nearly all circumstances. To avoid this assumption, we propose a solution based on a simple sample-split approach, that requires a much more realistic weak dependence condition instead. To acknowledge that the conditional quantities vary over time, we employ a merging concept generalizing the notion of weak convergence. The conditional distributions of the sample-split estimator and the estimator based on the two-independent-processes assumption are shown to merge in probability under mild conditions. The corresponding sample-split intervals are shown to coincide asymptotically with the intervals commonly constructed by practitioners, which provides a novel, and realistic, justification for confidence intervals for conditional objects, applicable to many common time series models such as ARMA and (extensions of) GARCH models.

One limitation to our approach is that we restrict ourselves to univariate time series and objects of interests. At the expense of more involved notation this could be readily extended to multivariate time series and objects of interests. A second, and more restrictive, limitation is our weak dependence assumption needed to achieve asymptotic independence between the two subsamples, which for instance rules out application to integrated processes. Given the fundamental role of this assumption in our setup, it appears difficult to generalize this. However, this also casts further doubt on the two-independent-processes assumption as validation for confidence intervals constructed for such persistent processes. A case-by-case treatment, as for instance done by Gospodinov (2002) for near unit root processes and Samaranayake and Hasza (1988) for explosive processes, appears to be necessary in such cases, and standard

confidence intervals should be treated with caution.

## Acknowledgements

The authors thank participants of the GSBE Econometric Seminar in Maastricht 2017 and the 1<sup>st</sup> International Conference on Econometrics and Statistics in Hong Kong 2017. Furthermore, the authors thank Franz Palm, Lenard Lieb, Denis de Crombrughe and Hanno Reuvers for useful suggestions and comments on earlier versions of the paper. This research was financially supported by the Netherlands Organisation for Scientific Research (NWO).

## A Proofs

We only present the proofs of the leading results here. The proofs of the remaining lemmas and corollaries can be found in Appendix B. Before going to the proofs, we first introduce the following auxiliary metrics that will be encountered in the proofs. For arbitrary cdfs  $F$  and  $G$  on  $\mathbb{R}$ , the Kolmogorov and Lévy metric are given by

$$d_K(F, G) = \sup_{\tau \in \mathbb{R}} |F(\tau) - G(\tau)|$$

$$d_L(F, G) = \inf \left\{ \xi > 0 : G(\tau - \xi) - \xi \leq F(\tau) \leq G(\tau + \xi) + \xi \quad \forall \tau \in \mathbb{R} \right\}.$$

Moreover, let  $Z \sim P_{\theta_0, \xi_0}$  be defined on some probability space  $(\check{\Omega}, \check{\mathcal{F}}, \check{\mathbb{P}})$  and define the product measure  $\bar{\mathbb{P}} = \mathbb{P} \times \check{\mathbb{P}}$  on the space  $\Omega \times \check{\Omega}$  with  $\sigma$ -field, generated by the *measurable rectangles*  $A \times \check{A}$  with  $A \in \mathcal{F}$  and  $\check{A} \in \check{\mathcal{F}}$  (cf. Billingsley, 1986, Thm. 18.2). Notice that  $Z$  is independent of  $\{X_t\}$ ,  $\{Y_t\}$  and  $S_0$  by construction.

**Lemma 1.** *Let  $R_n^{2IP}$  and  $R_n^{SPL}$  be the remainder terms as given in (4.4) and (4.5).*

(i) If Assumptions 1.1, 1.2(a), 1.3 – 1.5 hold, then  $R_n^{2IP}$  is  $o_p(1)$ ;

(ii) if Assumptions 2.1, 2.3 – 2.5 hold, then  $R_n^{SPL}$  is  $o_p(1)$ .

**Proof.** See Appendix B. □

**Lemma 2.** Let  $P_{Z_n}^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n)$  be as given in Assumption 2.2 and denote the conditional law of  $Z_n^{2IP} = m_n(\hat{\theta}(\mathbf{Y}_n) - \theta_0)$  given  $\mathcal{G}_1^n$  by  $P_{Z_n}^{2IP}(\cdot | \mathcal{G}_1^n)$ .

(i) Under Assumption 1.1 and 1.2,  $\int h dP_{Z_n}^{2IP}(\cdot | \mathcal{G}_1^n) \xrightarrow{p} \int h dP_{\theta_0, \xi_0}$  for all  $h \in BL$ ;

(ii) Under Assumption 2.1 and 2.2,  $\int h dP_{Z_n}^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n) \xrightarrow{p} \int h dP_{\theta_0, \xi_0}$  for all  $h \in BL$ .

**Proof.** See Appendix B. □

**Lemma 3.** Let  $Q_n$  be a sequence of regular conditional distributions and  $Q$  be a (non-random) law on  $(\mathbb{R}^r, \|\cdot\|)$ . If  $\int h dQ_n \xrightarrow{p} \int h dQ$  for all  $h \in BL$ , then  $\sup_{h \in \mathcal{H}} \left| \int h d(Q_n - Q) \right| \xrightarrow{p} 0$ , where  $\mathcal{H} = \{h : \mathbb{R}^r \rightarrow \mathbb{R} : \|h\|_{BL} \leq 1\}$ .

**Proof.** For  $r = 1$  Lemma 3 appears as Lemma 2 of the supplemental material to Castillo and Rousseau (2015). See Appendix B for a different proof, extending their result to  $r > 1$ . □

**Lemma 4.** Assume that the  $\mathbb{R}^r$ -valued random variable  $w_n$  is  $O_p(1)$  and  $\mathcal{F}_n$ -measurable. Further, suppose the real-valued random variable  $R_n$  is  $o_p(1)$  and the  $\mathbb{R}^r$ -valued random variable  $Z_n$  satisfies  $\bar{\mathbb{P}}[Z_n \in \cdot | \mathcal{F}_n] \xrightarrow{p} P_{\theta_0, \xi_0}$ . Then, the two regular conditional distributions  $\bar{\mathbb{P}}[w'_n Z_n + R_n \in \cdot | \mathcal{F}_n]$  and  $\bar{\mathbb{P}}[w'_n Z \in \cdot | \mathcal{F}_n]$  merge in probability.

**Proof.** This lemma is related to Lemma 8 in Belyaev and Sjöstedt-De Luna (2000) where the quantity corresponding to  $\bar{\mathbb{P}}[w'_n Z \in \cdot | \mathcal{F}_n]$  is non-random.



Set  $\mathcal{F} = \{f : \mathbb{R} \rightarrow \mathbb{R} : \|f\|_{BL} \leq 1\}$ . The triangle inequality implies

$$\begin{aligned} & \sup_{f \in \mathcal{F}} \left| \int [f(w'_n Z_n + R_n) - f(w'_n Z)] d\bar{\mathbb{P}}[\cdot | \mathcal{F}_n] \right| \\ & \leq \underbrace{\sup_{f \in \mathcal{F}} \left| \int [f(w'_n Z_n + R_n) - f(w'_n Z_n)] d\bar{\mathbb{P}}[\cdot | \mathcal{F}_n] \right|}_{=I} \\ & \quad + \underbrace{\sup_{f \in \mathcal{F}} \left| \int [f(w'_n Z_n) - f(w'_n Z)] d\bar{\mathbb{P}}[\cdot | \mathcal{F}_n] \right|}_{=II}. \end{aligned}$$

We show that  $I \xrightarrow{P} 0$  and  $II \xrightarrow{P} 0$ . Let  $\epsilon > 0$ ; as  $\|f\|_{BL} \leq 1$  for all  $f \in \mathcal{F}$  we obtain

$$\begin{aligned} I & \leq \sup_{f \in \mathcal{F}} \int |f(w'_n Z_n + R_n) - f(w'_n Z_n)| d\mathbb{P}[\cdot | \mathcal{F}_n] \\ & = \sup_{f \in \mathcal{F}} \int_{|R_n| \leq \epsilon} |f(w'_n Z_n + R_n) - f(w'_n Z_n)| d\bar{\mathbb{P}}[\cdot | \mathcal{F}_n] \\ & \quad + \sup_{f \in \mathcal{F}} \int_{|R_n| > \epsilon} |f(w'_n Z_n + R_n) - f(w'_n Z_n)| d\bar{\mathbb{P}}[\cdot | \mathcal{F}_n] \\ & \leq \sup_{f \in \mathcal{F}} \int_{|R_n| \leq \epsilon} \|f\|_{BL} |w'_n Z_n + R_n - w'_n Z_n| d\bar{\mathbb{P}}[\cdot | \mathcal{F}_n] \\ & \quad + \sup_{f \in \mathcal{F}} \int_{|R_n| > \epsilon} (|f(w'_n Z_n + R_n)| + |f(w'_n Z_n)|) d\bar{\mathbb{P}}[\cdot | \mathcal{F}_n] \\ & \leq \sup_{f \in \mathcal{F}} \|f\|_{BL} \int_{|R_n| \leq \epsilon} |R_n| d\mathbb{P}[\cdot | \mathcal{F}_n] + 2 \sup_{f \in \mathcal{F}} \|f\|_{BL} \int_{|R_n| > \epsilon} d\bar{\mathbb{P}}[\cdot | \mathcal{F}_n] \\ & \leq \int_{|R_n| \leq \epsilon} \epsilon d\bar{\mathbb{P}}[\cdot | \mathcal{F}_n] + 2 \bar{\mathbb{P}}[|R_n| > \epsilon | \mathcal{F}_n] \leq \epsilon + 2 \bar{\mathbb{P}}[|R_n| > \epsilon | \mathcal{F}_n]. \end{aligned}$$

In line with Xiong and Li (2008, Thm. 3.3), employing the Markov inequality we have

$$\bar{\mathbb{P}}[I \geq 2\epsilon] \leq \bar{\mathbb{P}}[\bar{\mathbb{P}}[|R_n| > \epsilon | \mathcal{F}_n] \geq \epsilon/2] \leq \frac{2}{\epsilon} \bar{\mathbb{P}}[|R_n| > \epsilon] \rightarrow 0$$

as  $R_n = o_p(1)$  and hence  $I = o_p(1)$ . Consider  $II$  and let  $K \geq 1$ . We obtain

$$\bar{\mathbb{P}}[II \geq \epsilon] \leq \bar{\mathbb{P}}[\|w_n\| \geq K] + \bar{\mathbb{P}}[II \geq \epsilon \cap \|w_n\| \leq K] .$$

As  $\|w_n\| = O_p(1)$  the first term can be made arbitrarily small by choosing  $K$  large.

For such  $K$ , consider the second term and note that

$$\begin{aligned} \bar{\mathbb{P}}[II \geq \epsilon \cap \|w_n\| \leq K] &= \bar{\mathbb{P}}\left[\sup_{f \in \mathcal{F}} \left| \int [f(w'_n Z_n) - f(w'_n Z)] d\bar{\mathbb{P}}[\cdot | \mathcal{F}_n] \right| \geq \epsilon \cap \|w_n\| \leq K\right] \\ &\leq \bar{\mathbb{P}}\left[\sup_{g \in \mathcal{G}} \left| \int [g(Z_n) - g(Z)] d\bar{\mathbb{P}}[\cdot | \mathcal{F}_n] \right| \geq \epsilon \cap \|w_n\| \leq K\right] \\ &\leq \bar{\mathbb{P}}\left[\sup_{g \in \mathcal{G}} \left| \int [g(Z_n) - g(Z)] d\bar{\mathbb{P}}[\cdot | \mathcal{F}_n] \right| \geq \epsilon\right] \\ &= \bar{\mathbb{P}}\left[\sup_{g \in \mathcal{G}} \left| \int g d\left(\mathbb{P}[Z_n \in \cdot | \mathcal{F}_n] - P_{\theta_0, \xi_0}\right) \right| \geq \epsilon\right] , \end{aligned}$$

where  $\mathcal{G} = \{g : \mathbb{R}^r \rightarrow \mathbb{R} \mid g(x) = f(w'x), \text{ for some } f \in \mathcal{F} \text{ and some } w \in \mathbb{R}^r \text{ with } \|w\| \leq K\}$ . We have that  $\|\cdot\|_{BL}$  is uniformly bounded for  $\mathcal{G}$  since for every  $g \in \mathcal{G}$

$$\begin{aligned} \|g\|_{BL} &= \sup_x |f(w'x)| + \sup_{x \neq y} \frac{|f(w'x) - f(w'y)|}{|w'x - w'y|} \frac{|w'x - w'y|}{\|x - y\|} \\ &\leq \sup_x |f(w'x)| + \sup_{x \neq y} \frac{|f(w'x) - f(w'y)|}{|w'x - w'y|} \|w\| \leq \|f\|_{BL} K \leq K . \end{aligned}$$

Thus,  $\|g/K\|_{BL} \leq 1$  and it follows by  $\bar{\mathbb{P}}[Z_n \in \cdot | \mathcal{F}_n] \xrightarrow{P} P_{\theta_0, \xi_0}$  and Lemma 3 that

$$\begin{aligned} \bar{\mathbb{P}}[II \geq \epsilon \cap \|w_n\| \leq K] &\leq \bar{\mathbb{P}}\left[\sup_{g \in \mathcal{G}} \left| \int \frac{g}{K} d\left(\mathbb{P}[Z_n \in \cdot | \mathcal{F}_n] - P_{\theta_0, \xi_0}\right) \right| \geq \frac{\epsilon}{K}\right] \\ &\leq \bar{\mathbb{P}}\left[\sup_{h \in \mathcal{H}} \left| \int h d\left(\mathbb{P}[Z_n \in \cdot | \mathcal{F}_n] - P_{\theta_0, \xi_0}\right) \right| \geq \frac{\epsilon}{K}\right] \rightarrow 0 , \end{aligned}$$

where  $\mathcal{H}$  is defined in Lemma 3. Thus,  $II$  is  $o_p(1)$ , which completes the proof.  $\square$

**Proof of Theorem 1.** Let  $w_n = \frac{\partial \psi_n(s_0^\circ, \mathbf{X}_n; \theta_0)}{\partial \theta}$  and  $v_n = \frac{\partial \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \theta_0)}{\partial \theta_0}$  and set

$$\bar{P}_n^{2IP}(\cdot | \mathcal{G}_1^n) = \bar{\mathbb{P}}[w'_n Z \in \cdot | \mathcal{G}_1^n] \quad (\text{A.1})$$

$$\bar{P}_n^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n) = \bar{\mathbb{P}}[v'_n Z \in \cdot | \mathcal{G}_{n-n_2+1}^n], \quad (\text{A.2})$$

the regular conditional distribution of  $w'_n Z$  given  $\mathcal{G}_1^n$  and  $v'_n Z$  given  $\mathcal{G}_{n-n_2+1}^n$ , respectively. The triangle inequality implies

$$\begin{aligned} d_{BL}\left(P_n^{2IP}(\cdot | \mathcal{G}_1^n), P_n^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n)\right) &\leq \underbrace{d_{BL}\left(P_n^{2IP}(\cdot | \mathcal{G}_1^n), \bar{P}_n^{2IP}(\cdot | \mathcal{G}_1^n)\right)}_I \\ &\quad + \underbrace{d_{BL}\left(\bar{P}_n^{2IP}(\cdot | \mathcal{G}_1^n), \bar{P}_n^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n)\right)}_{II} + \underbrace{d_{BL}\left(\bar{P}_n^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n), P_n^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n)\right)}_{III}. \end{aligned}$$

We prove that  $I$ ,  $II$  and  $III$  are  $o_p(1)$ . Consider  $I$  and note that  $m_n(\hat{\lambda}_{n+1}^{2IP} - \lambda_{n+1}) = w'_n Z_n^{2IP} + R_n^{2IP}$ , where  $Z_n^{2IP}$  is defined in Lemma 2.  $w_n$  is  $\mathcal{G}_1^n$  measurable and  $O_p(1)$  by Assumption 1.6,  $R_n^{2IP}$  is  $o_p(1)$  by Lemma 1 and  $\int h dP_{Z_n}^{2IP}(\cdot | \mathcal{G}_1^n) \xrightarrow{p} \int h dP_{\theta_0, \xi_0}$  for each  $h \in BL$  by Lemma 2. Replacing  $Z_n$ ,  $R_n$ ,  $w_n$  and  $\mathcal{F}_n$  in Lemma 4 by  $Z_n^{2IP}$ ,  $R_n^{2IP}$ ,  $w_n$  and  $\mathcal{G}_1^n$  implies that  $P_n^{2IP}(\cdot | \mathcal{G}_1^n) = \bar{\mathbb{P}}[w'_n Z_n^{2IP} + R_n^{2IP} \in \cdot | \mathcal{G}_1^n]$  and  $\bar{P}_n^{2IP}(\cdot | \mathcal{G}_1^n) = \bar{\mathbb{P}}[w'_n Z \in \cdot | \mathcal{G}_1^n]$  merge in probability, i.e.  $I \xrightarrow{p} 0$ .

Consider  $II$  and rewrite  $v'_n Z = w'_n Z + (v_n - w_n)'Z$ .  $w_n$  is  $\mathcal{G}_1^n$  measurable and  $O_p(1)$  by Assumption 1.6,  $(v_n - w_n)'Z = o_p(1) O_p(1) = o_p(1)$  by Assumption 3 and  $\int h d\bar{P}[Z \in \cdot | \mathcal{G}_1^n] = \int h dP_{\theta_0, \xi_0}$  for each  $h \in BL$ . Replacing  $Z_n$ ,  $R_n$ ,  $w_n$  and  $\mathcal{F}_n$  in Lemma 4 by  $Z$ ,  $(v_n - w_n)'Z$ ,  $w_n$  and  $\mathcal{G}_1^n$  implies that  $\bar{P}_n^{2IP}(\cdot | \mathcal{G}_1^n) = \bar{\mathbb{P}}[w'_n Z \in \cdot | \mathcal{G}_1^n]$  and  $\bar{P}_n^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n) = \bar{\mathbb{P}}[v'_n Z \in \cdot | \mathcal{G}_1^n]$  merge in probability, i.e.  $II \xrightarrow{p} 0$ .

Consider  $III$  and note that  $m_n(\hat{\lambda}_{n+1}^{SPL} - \lambda_{n+1}) = v'_n Z_n^{SPL} + R_n^{SPL}$ , where  $Z_n^{SPL}$  is defined in Lemma 2.  $v_n$  is  $\mathcal{G}_{n-n_2+1}^n$  measurable and  $O_p(1)$  by Assumption 2.6,  $R_n^{SPL}$  is  $o_p(1)$  by Lemma 1 and  $\int h dP_{Z_n}^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n) \xrightarrow{p} \int h dP_{\theta_0, \xi_0}$  for each  $h \in BL$  by Lemma

2. Replacing  $Z_n$ ,  $R_n$ ,  $w_n$  and  $\mathcal{F}_n$  in Lemma 4 by  $Z_n^{SPL}$ ,  $R_n^{SPL}$ ,  $v_n$  and  $\mathcal{G}_{n-n_2+1}^n$  implies that  $P_n^{SPL}(\cdot|\mathcal{G}_{n-n_2+1}^n) = \bar{\mathbb{P}}[v_n' Z_n^{SPL} + R_n^{SPL} \in \cdot|\mathcal{G}_{n-n_2+1}^n]$  and  $\bar{\mathbb{P}}[v_n' Z \in \cdot|\mathcal{G}_{n-n_2+1}^n] = \bar{P}_n^{SPL}(\cdot|\mathcal{G}_{n-n_2+1}^n)$  merge in probability, i.e.  $III \xrightarrow{p} 0$ .  $\square$

**Lemma 5.** *Let  $\epsilon > 0$  and  $F$  and  $G$  be cdfs on  $\mathbb{R}$  with  $G(\tau-\epsilon)-\epsilon \leq F(\tau) \leq G(\tau+\epsilon)+\epsilon$  for all  $\tau \in \mathbb{R}$ . Then  $F^{-1}(u-\epsilon)-\epsilon \leq G^{-1}(u) \leq F^{-1}(u+\epsilon)+\epsilon$  for all  $u \in (\epsilon, 1-\epsilon)$ .*

**Proof.** See Appendix B.  $\square$

**Lemma 6.** *Suppose  $F_n$  and  $G_n$  are sequences of conditional cdfs with  $d_L(F_n, G_n) \xrightarrow{p} 0$  as  $n \rightarrow \infty$ . Further, assume that  $G_n$  is stochastically uniformly equicontinuous: for every  $\epsilon, \eta > 0$ , there exist  $\delta = \delta(\epsilon, \eta) > 0$  and  $N = N(\epsilon, \eta)$  such that  $\mathbb{P}\left[\sup_{\tau \in \mathbb{R}} \sup_{\tau' \in \mathbb{R}: |\tau-\tau'| < \delta} |G_n(\tau') - G_n(\tau)| > \epsilon\right] < \eta$  for all  $n \geq N$ . Then,  $d_K(F_n, G_n) \xrightarrow{p} 0$ .*

**Proof.** This lemma is related to Lemma 9 in Belyaev and Sjöstedt-De Luna (2000) where the sequence  $G_n$  is non-random. See Appendix B for the proof.  $\square$

**Lemma 7.** *Let  $P_n$  and  $Q_n$  be sequences of regular conditional distributions with corresponding cdf  $F_n$  and  $G_n$ , respectively. If  $P_n$  and  $Q_n$  merge in probability and  $G_n$  is stochastically uniformly equicontinuous, then  $\mathbb{P}[G_n^{-1}(\gamma_1) \leq T_n \leq G_n^{-1}(1-\gamma_2)|\mathcal{F}_n] \xrightarrow{p} 1-\gamma$ , where  $\gamma = \gamma_1 + \gamma_2$  and  $T_n$  given the  $\sigma$ -algebra  $\mathcal{F}_n$  follows the law  $P_n$ .*

**Proof.** Since  $P_n$  and  $Q_n$  merge in probability and  $d_L \leq 2d_{BL}^{1/2}$  (cf. Huber, 2009, p. 36; Dudley, 2002, Thm. 11.3.3), we have  $d_L(F_n, G_n) \xrightarrow{p} 0$ . Let  $u \in (0, 1)$  and take  $\epsilon > 0$  sufficiently small satisfying  $u \in (\epsilon, 1-\epsilon)$ .  $\mathbb{P}[d_L(F_n, G_n) > \epsilon|\mathcal{F}_n]$  is  $o_p(1)$  since for every  $\delta > 0$  the Markov inequality implies  $\mathbb{P}\left[\mathbb{P}[d_L(F_n, G_n) > \epsilon|\mathcal{F}_n] \geq \delta\right] \leq$

$\frac{1}{\delta} \mathbb{P}[d_L(F_n, G_n) > \epsilon] \rightarrow 0$ . Employing Lemma 5 we derive the following bounds:

$$\begin{aligned}
& \mathbb{P}[T_n \leq G_n^{-1}(u) | \mathcal{F}_n] \\
& \leq \mathbb{P}[T_n \leq G_n^{-1}(u) \cap d_L(F_n, G_n) \leq \epsilon | \mathcal{F}_n] + \mathbb{P}[d_L(F_n, G_n) > \epsilon | \mathcal{F}_n] \\
& \leq \mathbb{P}[T_n \leq F_n^{-1}(u + \epsilon) + \epsilon \cap d_L(F_n, G_n) \leq \epsilon | \mathcal{F}_n] + o_p(1) \\
& \leq \mathbb{P}[T_n \leq F_n^{-1}(u + \epsilon) + \epsilon | \mathcal{F}_n] + o_p(1) \\
& = F_n(F_n^{-1}(u + \epsilon) + \epsilon) + o_p(1) = U_n
\end{aligned}$$

$$\begin{aligned}
& \mathbb{P}[T_n < G_n^{-1}(u) | \mathcal{F}_n] \\
& \geq \mathbb{P}[T_n < G_n^{-1}(u) \cap d_L(F_n, G_n) \leq \epsilon | \mathcal{F}_n] \\
& \geq \mathbb{P}[T_n < F_n^{-1}(u - \epsilon) - \epsilon \cap d_L(F_n, G_n) \leq \epsilon | \mathcal{F}_n] \\
& \geq \mathbb{P}[T_n < F_n^{-1}(u - \epsilon) - \epsilon | \mathcal{F}_n] - \mathbb{P}[d_L(F_n, G_n) > \epsilon | \mathcal{F}_n] \\
& = F_n(F_n^{-1}(u - \epsilon) - \epsilon - ) - o_p(1) = L_n,
\end{aligned}$$

where  $F_n(\cdot -)$  denotes the left limit of  $F_n(\cdot)$ . We show that  $L_n$  and  $U_n$  converge in probability to  $u$ . Regarding the lower bound  $L_n$  we have

$$\begin{aligned}
& \left| F_n(F_n^{-1}(u - \epsilon) - \epsilon -) - u \right| \\
& \leq \left| F_n(F_n^{-1}(u - \epsilon) - \epsilon -) - F_n(F_n^{-1}(u - \epsilon) -) \right| + \left| F_n(F_n^{-1}(u - \epsilon) -) - (u - \epsilon) \right| + \epsilon \\
& \leq \sup_{\tau \in \mathbb{R}} \left| F_n(\tau - \epsilon -) - F_n(\tau -) \right| + \left| F_n(F_n^{-1}(u - \epsilon) -) - (u - \epsilon) \right| + \epsilon \\
& \leq \sup_{\tau \in \mathbb{R}} \left| F_n(\tau - \epsilon -) - F_n(\tau -) \right| + \sup_{\tau \in \mathbb{R}} \left| F_n(\tau) - F_n(\tau -) \right| + \epsilon \\
& \leq 4d_K(F_n, G_n) + \sup_{\tau \in \mathbb{R}} \left| G_n(\tau - \epsilon -) - G_n(\tau -) \right| + \sup_{\tau \in \mathbb{R}} \left| G_n(\tau) - G_n(\tau -) \right| + \epsilon,
\end{aligned}$$

where the third inequality is due to Cavaliere et al. (2013, p. 217). As  $d_L(F_n, G_n) \xrightarrow{p} 0$

since  $d_L \leq 2d_{BL}^{1/2}$  (cf. Huber, 2009, p. 36; Dudley, 2002, Thm. 11.3.3) and  $G_n$  is stochastically uniformly equicontinuous, Lemma 6 implies  $d_K(F_n, G_n) \xrightarrow{P} 0$ . Further,  $\sup_{\tau \in \mathbb{R}} |G_n(\tau - \epsilon -) - G_n(\tau -)| = o_p(1)$  and  $\sup_{\tau \in \mathbb{R}} |G_n(\tau) - G_n(\tau -)| = o_p(1)$  by stochastic uniform equicontinuity completing  $L_n \xrightarrow{P} u$ . Regarding the upper bound  $U_n$  we have

$$\begin{aligned}
& \left| F_n(F_n^{-1}(u + \epsilon) + \epsilon) - u \right| \\
& \leq \left| F_n(F_n^{-1}(u + \epsilon) + \epsilon) - F_n(F_n^{-1}(u + \epsilon)) \right| + \left| F_n(F_n^{-1}(u + \epsilon)) - F_n(F_n^{-1}(u + \epsilon) -) \right| \\
& \quad + \left| F_n(F_n^{-1}(u + \epsilon) -) - (u + \epsilon) \right| + \epsilon \\
& \leq \sup_{\tau \in \mathbb{R}} \left| F_n(\tau + \epsilon) - F_n(\tau) \right| + 2 \sup_{\tau \in \mathbb{R}} \left| F_n(\tau) - F_n(\tau -) \right| + \epsilon \\
& \leq 6d_K(F_n, G_n) + \sup_{\tau \in \mathbb{R}} \left| G_n(\tau + \epsilon) - G_n(\tau) \right| + 2 \sup_{\tau \in \mathbb{R}} \left| G_n(\tau) - G_n(\tau -) \right| + \epsilon,
\end{aligned}$$

where all terms on the right hand side are  $o_p(1)$  such that  $U_n \xrightarrow{P} u$ . We obtain

$$\underbrace{L_n(u)}_{\xrightarrow{P} u} \leq \mathbb{P}[T_n < G_n^{-1}(u) | \mathcal{F}_n] \leq \mathbb{P}[T_n \leq G_n^{-1}(u) | \mathcal{F}_n] \leq \underbrace{U_n(u)}_{\xrightarrow{P} u},$$

which implies that  $\mathbb{P}[T_n < G_n^{-1}(u) | \mathcal{F}_n]$  and  $\mathbb{P}[T_n \leq G_n^{-1}(u) | \mathcal{F}_n]$  converge in probability to  $u$  for arbitrary  $u \in (0, 1)$ ; in particular  $\gamma_1$  and  $1 - \gamma_2$ . It follows that

$$\begin{aligned}
& \mathbb{P}[G_n^{-1}(\gamma_1) \leq T_n \leq G_n^{-1}(1 - \gamma_2) | \mathcal{F}_n] \\
& = \mathbb{P}[T_n \leq G_n^{-1}(1 - \gamma_2) | \mathcal{F}_n] - \mathbb{P}[T_n < G_n^{-1}(\gamma_1) | \mathcal{F}_n] \xrightarrow{P} 1 - \gamma_2 - \gamma_1 = 1 - \gamma,
\end{aligned}$$

which completes the proof.  $\square$

**Proof of Theorem 2.** Consider statement 1(a). Let  $\hat{Z}_{\hat{\theta}(\mathbf{Y}_n), \hat{\xi}(\mathbf{Y}_n)}$  follow the mixture distribution  $P_{\hat{\theta}(\mathbf{Y}_n), \hat{\xi}(\mathbf{Y}_n)}$  such that, given  $\hat{\theta}(\mathbf{Y}_n) = \bar{\theta}$  and  $\hat{\xi}(\mathbf{Y}_n) = \bar{\xi}$ , the conditional distribution of  $\hat{Z}_{\hat{\theta}(\mathbf{Y}_n), \hat{\xi}(\mathbf{Y}_n)}$  is the respective member of the parametric family  $\{P_{\theta, \eta} | \theta \in$

$\Theta, \xi \in \Xi$ , i.e. the conditional law is  $P_{\hat{\theta}, \hat{\xi}}$ . Further, let

$$\widehat{P_n^{2IP}}(\cdot) \text{ be the conditional law of } \hat{w}_n' \hat{Z}_{\hat{\theta}(\mathbf{Y}_n), \hat{\xi}(\mathbf{Y}_n)} \text{ given } \mathcal{H}_n \quad (\text{A.3})$$

where  $\hat{w}_n = \frac{\partial \psi_n(s_0^\circ, \mathbf{X}_n; \hat{\theta}(\mathbf{Y}_n))}{\partial \theta}$  and  $\mathcal{H}_n = \sigma(X_1, \dots, X_n, \hat{\theta}(\mathbf{Y}_n), \hat{\xi}(\mathbf{Y}_n))$ . The triangle inequality implies that

$$d_{BL}\left(P_n^{2IP}(\cdot|\mathcal{G}_1^n), \widehat{P_n^{2IP}}(\cdot)\right) \leq d_{BL}\left(P_n^{2IP}(\cdot|\mathcal{G}_1^n), \bar{P}_n^{2IP}(\cdot|\mathcal{G}_1^n)\right) + d_{BL}\left(\bar{P}_n^{2IP}(\cdot|\mathcal{G}_1^n), \widehat{P_n^{2IP}}(\cdot)\right),$$

where  $\bar{P}_n^{2IP}(\cdot|\mathcal{G}_1^n)$  is defined in equation (A.1). In the proof of Theorem 1, we have shown that  $P_n^{2IP}(\cdot|\mathcal{G}_1^n)$  and  $\bar{P}_n^{2IP}(\cdot|\mathcal{G}_1^n)$  merge in probability under Assumption 1. It suffices to show that  $\bar{P}_n^{2IP}(\cdot|\mathcal{G}_1^n)$  and  $\widehat{P_n^{2IP}}(\cdot)$  merge in probability. Write  $\hat{w}_n' \hat{Z}_{\hat{\theta}(\mathbf{Y}_n), \hat{\xi}(\mathbf{Y}_n)} = w_n' \hat{Z}_{\hat{\theta}(\mathbf{Y}_n), \hat{\xi}(\mathbf{Y}_n)} + \hat{R}_n^{2IP}$  with  $\hat{R}_n^{2IP} = (\hat{w}_n - w_n)' \hat{Z}_{\hat{\theta}(\mathbf{Y}_n), \hat{\xi}(\mathbf{Y}_n)}$ . First, we show  $\hat{R}_n^{2IP} = o_p(1)$ . Take an arbitrary  $\epsilon > 0$ . We obtain

$$\begin{aligned} \mathbb{P}\left[|\hat{R}_n^{2IP}| \geq \epsilon\right] &\leq \mathbb{P}\left[\left\|\frac{\partial^2 \psi_n(s_0^\circ, \mathbf{X}_n; \dot{\theta}_n)}{\partial \theta \partial \theta'}\right\| \left\|\hat{\theta}(\mathbf{Y}_n) - \theta_0\right\| \left\|\hat{Z}_{\hat{\theta}(\mathbf{Y}_n), \hat{\xi}(\mathbf{Y}_n)}\right\| \geq \epsilon\right] \\ &\leq \mathbb{P}\left[\left\|\frac{\partial^2 \psi_n(s_0^\circ, \mathbf{X}_n; \dot{\theta}_n)}{\partial \theta \partial \theta'}\right\| \left\|\hat{\theta}(\mathbf{Y}_n) - \theta_0\right\| \left\|\hat{Z}_{\hat{\theta}(\mathbf{Y}_n), \hat{\xi}(\mathbf{Y}_n)}\right\| \geq \epsilon \cap \dot{\theta}_n \in \Theta^*\right] + \mathbb{P}\left[\dot{\theta}_n \notin \Theta^*\right] \\ &\leq \mathbb{P}\left[\sup_{\theta \in \Theta^*} \left\|\frac{\partial^2 \psi_n(s_0^\circ, \mathbf{X}_n; \theta)}{\partial \theta \partial \theta'}\right\| \left\|\hat{\theta}(\mathbf{Y}_n) - \theta_0\right\| \left\|\hat{Z}_{\hat{\theta}(\mathbf{Y}_n), \hat{\xi}(\mathbf{Y}_n)}\right\| \geq \epsilon\right] + \mathbb{P}\left[\dot{\theta}_n \notin \Theta^*\right], \end{aligned}$$

where  $\dot{\theta}_n$  lies between  $\hat{\theta}(\mathbf{Y}_n)$  and  $\theta_0$ . The first term vanishes as  $\sup_{\theta \in \Theta^*} \left\|\frac{\partial^2 \psi_n(s_0^\circ, \mathbf{X}_n; \theta)}{\partial \theta \partial \theta'}\right\| = O_p(1)$  by Assumption 1.5,  $\left\|\hat{\theta}(\mathbf{Y}_n) - \theta_0\right\| = O_p(m_n^{-1}) = o_p(1)$  by Assumptions 1.1 and 1.2(a) and  $\left\|\hat{Z}_{\hat{\theta}(\mathbf{Y}_n), \hat{\xi}(\mathbf{Y}_n)}\right\| = O_p(1)$  as  $\hat{Z}_{\hat{\theta}(\mathbf{Y}_n), \hat{\xi}(\mathbf{Y}_n)} \sim P_{\hat{\theta}(\mathbf{Y}_n), \hat{\xi}(\mathbf{Y}_n)} \xrightarrow{p} P_{\theta_0, \xi_0}$  by Assumptions 1.2(a) and 4.1. Further, as  $\hat{\theta}(\mathbf{Y}_n) \xrightarrow{p} \theta_0 \in \Theta^*$  and  $\Theta^*$  is open, we have  $\mathbb{P}[\dot{\theta}_n \notin \Theta^*] \rightarrow 0$  and  $\hat{R}_n^{2IP} = o_p(1)$  follows. Moreover,  $w_n$  is  $\mathcal{H}_n$ -measurable and  $O_p(1)$  by Assumption 1.6 and  $\int h d\bar{\mathbb{P}}[\hat{Z}_{\hat{\theta}(\mathbf{Y}_n), \hat{\xi}(\mathbf{Y}_n)} \in \cdot | \mathcal{H}_n] = \int h dP_{\hat{\theta}(\mathbf{Y}_n), \hat{\xi}(\mathbf{Y}_n)} \xrightarrow{p} \int h dP_{\theta_0, \xi_0}$

for each  $h \in BL$ . Replacing  $Z_n$ ,  $R_n$ ,  $w_n$  and  $\mathcal{F}_n$  in Lemma 4 by  $\hat{Z}_{\hat{\theta}(\mathbf{Y}_n), \hat{\xi}(\mathbf{Y}_n)}$ ,  $\hat{R}_n^{2IP}$ ,  $w_n$  and  $\mathcal{H}_n$  implies that  $\widehat{P_n^{2IP}}(\cdot) = \bar{\mathbb{P}}[w'_n \hat{Z}_{\hat{\theta}(\mathbf{Y}_n), \hat{\xi}(\mathbf{Y}_n)} + \hat{R}_n^{2IP} \in \cdot | \mathcal{H}_n]$  and  $\bar{\mathbb{P}}[w'_n Z \in \cdot | \mathcal{H}_n] = \bar{P}_n^{2IP}(\cdot | \mathcal{G}_1^n)$  merge in probability.

Consider statement 1(b). As  $P_n^{2IP}(\cdot | \mathcal{G}_1^n)$  and  $\widehat{P_n^{2IP}}(\cdot)$  merge in probability and  $\widehat{F_n^{2IP}}(\cdot)$  is assumed to be stochastically uniformly continuous, Lemma 7 applies. Replacing  $P_n$ ,  $Q_n$ ,  $F_n$ ,  $G_n$ ,  $T_n$  and  $\mathcal{F}_n$  by  $P_n^{2IP}(\cdot | \mathcal{G}_1^n)$ ,  $\widehat{P_n^{2IP}}(\cdot)$ ,  $F_n^{2IP}(\cdot | \mathcal{G}_1^n)$ ,  $\widehat{F_n^{2IP}}(\cdot)$ ,  $m_n(\hat{\lambda}_{n+1}^{2IP} - \lambda_{n+1})$  and  $\mathcal{G}_1^n$ , respectively, it follows that

$$\begin{aligned} & \mathbb{P}\left[I_\gamma^{2IP}(\mathbf{Y}_n, \mathbf{X}_n) \ni \lambda_{n+1} \middle| \mathcal{G}_1^n\right] \\ &= \mathbb{P}\left[\widehat{F_n^{2IP}}^{-1}(\gamma_1) \leq m_n(\hat{\lambda}_{n+1}^{2IP} - \lambda_{n+1}) \leq \widehat{F_n^{2IP}}^{-1}(1 - \gamma_2) \middle| \mathcal{G}_1^n\right] \xrightarrow{P} 1 - \gamma. \end{aligned}$$

Claim 2(a) is similarly proven as 1(a). Let  $\hat{Z}_{\hat{\theta}(\mathbf{x}_n^1), \hat{\xi}(\mathbf{x}_n^1)}$  follow the mixture distribution  $P_{\hat{\theta}(\mathbf{x}_n^1), \hat{\xi}(\mathbf{x}_n^1)}$  such that, given  $\hat{\theta}(\mathbf{X}_n^1) = \bar{\theta}$  and  $\hat{\xi}(\mathbf{X}_n^1) = \bar{\xi}$ , the conditional distribution of  $\hat{Z}_{\hat{\theta}(\mathbf{x}_n^1), \hat{\xi}(\mathbf{x}_n^1)}$  is the respective member of the parametric family  $\{P_{\theta, \eta} | \theta \in \Theta, \eta \in \Xi\}$ , i.e. the conditional law is  $P_{\bar{\theta}, \bar{\xi}}$ . Further, let

$$\widehat{P_n^{SPL}}(\cdot) \text{ be the conditional law of } \hat{v}_n \hat{Z}_{\hat{\theta}(\mathbf{x}_n^1), \hat{\xi}(\mathbf{x}_n^1)} \text{ given } \mathcal{I}_n \quad (\text{A.4})$$

where  $\hat{v}_n = \frac{\partial \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \hat{\theta}(\mathbf{X}_n^1))}{\partial \theta}$  and  $\mathcal{I}_n = \sigma(X_{n-n_2+1}, \dots, X_n, \hat{\theta}(\mathbf{X}_n^1), \hat{\xi}(\mathbf{X}_n^1))$ . The triangle inequality implies that

$$\begin{aligned} & d_{BL}\left(P_n^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n), \widehat{P_n^{SPL}}(\cdot)\right) \\ & \leq d_{BL}\left(P_n^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n), \bar{P}_n^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n)\right) + d_{BL}\left(\bar{P}_n^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n), \widehat{P_n^{SPL}}(\cdot)\right), \end{aligned}$$

where  $\bar{P}_n^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n)$  is defined in equation (A.2). In the proof of Theorem 1, we have shown that  $P_n^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n)$  and  $\bar{P}_n^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n)$  merge in probability under



Assumption 2. It suffices to show that  $\bar{P}_n^{SPL}(\cdot|\mathcal{G}_{n-n_2+1}^n)$  and  $\widehat{\bar{P}}_n^{SPL}(\cdot)$  merge in probability. Write  $\hat{v}'_n \hat{Z}_{\hat{\theta}(\mathbf{X}_n^1), \hat{\xi}(\mathbf{X}_n^1)} = v'_n \hat{Z}_{\hat{\theta}(\mathbf{X}_n^1), \hat{\xi}(\mathbf{X}_n^1)} + \hat{R}_n^{SPL}$  with  $\hat{R}_n^{SPL} = (\hat{v}_n - v_n)' \hat{Z}_{\hat{\theta}(\mathbf{X}_n^1), \hat{\xi}(\mathbf{X}_n^1)}$ . First, we show  $\hat{R}_n^{SPL} = o_p(1)$ . Take an arbitrary  $\epsilon > 0$ . We obtain

$$\begin{aligned} & \mathbb{P}\left[|\hat{R}_n^{SPL}| \geq \epsilon\right] \\ & \leq \mathbb{P}\left[\left\|\frac{\partial^2 \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \ddot{\theta}_n)}{\partial \theta \partial \theta'}\right\| \left\|\hat{\theta}(\mathbf{X}_n^1) - \theta_0\right\| \left\|\hat{Z}_{\hat{\theta}(\mathbf{X}_n^1), \hat{\xi}(\mathbf{X}_n^1)}\right\| \geq \epsilon\right] \\ & \leq \mathbb{P}\left[\left\|\frac{\partial^2 \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \ddot{\theta}_n)}{\partial \theta \partial \theta'}\right\| \left\|\hat{\theta}(\mathbf{X}_n^1) - \theta_0\right\| \left\|\hat{Z}_{\hat{\theta}(\mathbf{X}_n^1), \hat{\xi}(\mathbf{X}_n^1)}\right\| \geq \epsilon \cap \ddot{\theta}_n \in \Theta^*\right] + \mathbb{P}\left[\ddot{\theta}_n \notin \Theta^*\right] \\ & \leq \mathbb{P}\left[\sup_{\theta \in \Theta^*} \left\|\frac{\partial^2 \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \theta)}{\partial \theta \partial \theta'}\right\| \left\|\hat{\theta}(\mathbf{X}_n^1) - \theta_0\right\| \left\|\hat{Z}_{\hat{\theta}(\mathbf{X}_n^1), \hat{\xi}(\mathbf{X}_n^1)}\right\| \geq \epsilon\right] + \mathbb{P}\left[\ddot{\theta}_n \notin \Theta^*\right], \end{aligned}$$

where  $\ddot{\theta}_n$  lies between  $\hat{\theta}(\mathbf{X}_n^1)$  and  $\theta_0$ . The first term vanishes as  $\sup_{\theta \in \Theta^*} \left\|\frac{\partial^2 \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \theta)}{\partial \theta \partial \theta'}\right\|$  is  $O_p(1)$  by Assumption 2.5 and  $\left\|\hat{\theta}(\mathbf{X}_n^1) - \theta_0\right\| = O_p(m_n^{-1}) = o_p(1)$  by Assumptions 2.1,  $\left\|\hat{Z}_{\hat{\theta}(\mathbf{X}_n^1), \hat{\xi}(\mathbf{X}_n^1)}\right\| = O_p(1)$  as  $\hat{Z}_{\hat{\theta}(\mathbf{X}_n^1), \hat{\xi}(\mathbf{X}_n^1)} \sim P_{\hat{\theta}(\mathbf{X}_n^1), \hat{\xi}(\mathbf{X}_n^1)} \xrightarrow{p} P_{\theta_0, \xi_0}$  by Assumption 4.2. Further, as  $\hat{\theta}(\mathbf{X}_n^1) \xrightarrow{p} \theta_0 \in \Theta^*$  and  $\Theta^*$  is open, we have  $\mathbb{P}[\ddot{\theta}_n \notin \Theta^*] \rightarrow 0$  and  $\hat{R}_n^{SPL} = o_p(1)$  follows. Moreover,  $v_n$  is  $\mathcal{I}_n$ -measurable and  $O_p(1)$  by Assumption 2.6 and  $\int h d\bar{\mathbb{P}}[\hat{Z}_{\hat{\theta}(\mathbf{X}_n^1), \hat{\xi}(\mathbf{X}_n^1)} \in \cdot | \mathcal{I}_n] = \int h dP_{\hat{\theta}(\mathbf{X}_n^1), \hat{\xi}(\mathbf{X}_n^1)} \xrightarrow{p} \int h dP_{\theta_0, \xi_0}$  for each  $h \in BL$ . Replacing  $Z_n$ ,  $R_n$ ,  $w_n$  and  $\mathcal{F}_n$  in Lemma 4 by  $\hat{Z}_{\hat{\theta}(\mathbf{X}_n^1), \hat{\xi}(\mathbf{X}_n^1)}$ ,  $\hat{R}_n^{SPL}$ ,  $v_n$  and  $\mathcal{I}_n$  implies that  $\widehat{\bar{P}}_n^{SPL}(\cdot) = \bar{\mathbb{P}}[v'_n \hat{Z}_{\hat{\theta}(\mathbf{X}_n^1), \hat{\xi}(\mathbf{X}_n^1)} + \hat{R}_n^{SPL} \in \cdot | \mathcal{I}_n]$  and  $\bar{\mathbb{P}}[v'_n Z \in \cdot | \mathcal{I}_n] = \bar{P}_n^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n)$  merge in probability.

The proof of statement 2(b) is similar to the proof of claim 1(b). As  $P_n^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n)$  and  $\widehat{\bar{P}}_n^{SPL}(\cdot)$  merge in probability and  $\widehat{F}_n^{SPL}(\cdot)$  is assumed to be stochastically uniformly continuous, Lemma 7 applies. Replacing  $P_n$ ,  $Q_n$ ,  $F_n$ ,  $G_n$ ,  $T_n$  and  $\mathcal{F}_n$  by  $P_n^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n)$ ,  $\widehat{\bar{P}}_n^{SPL}(\cdot)$ ,  $F_n^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n)$ ,  $\widehat{F}_n^{SPL}(\cdot)$ ,  $m_n(\hat{\lambda}_{n+1}^{SPL} - \lambda_{n+1})$  and  $\mathcal{G}_{n-n_2+1}^n$ ,

respectively, it follows that

$$\begin{aligned} & \mathbb{P}\left[I_{\gamma}^{SPL}(\mathbf{X}_n^1, \mathbf{X}_n^2) \ni \lambda_{n+1} \middle| \mathcal{G}_{n-n_2+1}^n\right] \\ &= \mathbb{P}\left[\widehat{F}_n^{SPL^{-1}}(\gamma_1) \leq m_n(\hat{\lambda}_{n+1}^{2IP} - \lambda_{n+1}) \leq \widehat{F}_n^{SPL^{-1}}(1 - \gamma_2) \middle| \mathcal{G}_{n-n_2+1}^n\right] \end{aligned}$$

converges to  $1 - \gamma$  in probability, which completes the proof.  $\square$

**Lemma 8.** *Suppose  $F_n$  and  $G_n$  are sequences of conditional cdfs with  $d_L(F_n, G_n) \xrightarrow{p} 0$  as  $n \rightarrow \infty$ . Further, assume that  $G_n^{-1}$  is stochastically pointwise equicontinuous at  $u \in (0, 1)$ : for all  $\epsilon, \eta > 0$ , there exist  $\delta = \delta(\epsilon, \eta, u) > 0$  and  $N = N(\epsilon, \eta, u)$  such that  $\mathbb{P}\left[\sup_{|u-v| < \delta} |G_n^{-1}(v) - G_n^{-1}(u)| > \epsilon\right] < \eta$  for all  $n \geq N$ . Then  $|F_n^{-1}(u) - G_n^{-1}(u)| \xrightarrow{p} 0$ .*

**Proof.** See Appendix B.  $\square$

**Proof of Theorem 3.** Consider the first statement and write  $\hat{\lambda}_{n+1}^{SPL} - \hat{\lambda}_{n+1}^{ILL} = (\hat{\lambda}_{n+1}^{SPL} - \lambda_{n+1}) - (\hat{\lambda}_{n+1}^{ILL} - \lambda_{n+1})$ . We show that both terms are  $o_p(1)$ . Using (4.5), we have

$$\hat{\lambda}_{n+1}^{SPL} - \lambda_{n+1} = \frac{\partial \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \theta_0)}{\partial \theta'} (\hat{\theta}(\mathbf{X}_n^1) - \theta_0) + m_n^{-1} R_n^{SPL},$$

where  $\frac{\partial \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \theta_0)}{\partial \theta'} = O_p(1)$  by Assumption 2.6 and  $\hat{\theta}(\mathbf{X}_n^1) - \theta_0 = o_p(1)$  by Assumption 2.1. Together with  $R_n^{SPL} = o_p(1)$  by Lemma 1 and  $m_n^{-1} = o(1)$ , it implies that  $\hat{\lambda}_{n+1}^{SPL} - \lambda_{n+1} = o_p(1)$ . In addition, replacing  $\mathbf{Y}_n$  by  $\mathbf{X}_n$  in equation (4.4), we get

$$\hat{\lambda}_{n+1}^{ILL} - \lambda_{n+1} = \frac{\partial \psi_n(s_0^\circ, \mathbf{X}_n; \theta_0)}{\partial \theta'} (\hat{\theta}(\mathbf{X}_n) - \theta_0) + m_n^{-1} R_n^{ILL},$$

where  $R_n^{ILL}$  is obtained by replacing  $\mathbf{Y}_n$  by  $\mathbf{X}_n$  in  $R_n^{2IP}$ . We have  $\frac{\partial \psi_n(s_0^\circ, \mathbf{X}_n; \theta_0)}{\partial \theta'} = O_p(1)$  by Assumption 1.6 and  $\hat{\theta}(\mathbf{X}_n) - \theta_0 = o_p(1)$  by Assumption 1.1. Since  $R_n^{2IP} = o_p(1)$  has been shown in Lemma 1 without using Assumption 1.2(b), we have  $R_n^{ILL} = o_p(1)$ . Together with  $m_n^{-1} = o(1)$ , it follows that  $\hat{\lambda}_{n+1}^{ILL} - \lambda_{n+1} = o_p(1)$  completing the claim.

Consider the second statement. Let  $\hat{Z}_{\hat{\theta}(\mathbf{X}_n), \hat{\xi}(\mathbf{X}_n)}$  follow the mixture distribution  $P_{\hat{\theta}(\mathbf{X}_n), \hat{\xi}(\mathbf{X}_n)}$  such that, given  $\hat{\theta}(\mathbf{X}_n) = \bar{\theta}$  and  $\hat{\xi}(\mathbf{X}_n) = \bar{\xi}$ , the conditional law of  $\hat{Z}_{\hat{\theta}(\mathbf{X}_n), \hat{\xi}(\mathbf{X}_n)}$  is the respective member of the parametric family  $\{P_{\theta, \eta} | \theta \in \Theta, \eta \in \Xi\}$ , i.e. the conditional law is  $P_{\bar{\theta}, \bar{\xi}}$ . Further, let

$$\widehat{P}_n^{ILL}(\cdot) \text{ be the conditional law of } \hat{u}'_n \hat{Z}_{\hat{\theta}(\mathbf{X}_n), \hat{\xi}(\mathbf{X}_n)} \text{ given } \mathcal{F}_1^n, \quad (\text{A.5})$$

where  $\hat{u}_n = \frac{\partial \psi_n(s_0^\circ, \mathbf{X}_n; \hat{\theta}(\mathbf{X}_n))}{\partial \theta}$ . First, we show that  $\widehat{P}_n^{ILL}(\cdot)$  and  $\widehat{P}_n^{SPL}(\cdot)$ , defined in (A.4), merge in probability. The triangle inequality implies

$$\begin{aligned} d_{BL}\left(\widehat{P}_n^{SPL}(\cdot), \widehat{P}_n^{ILL}(\cdot)\right) &\leq d_{BL}\left(\widehat{P}_n^{SPL}(\cdot), P_n^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n)\right) \\ &\quad + d_{BL}\left(P_n^{SPL}(\cdot | \mathcal{G}_{n-n_2+1}^n), P_n^{2IP}(\cdot | \mathcal{G}_1^n)\right) + d_{BL}\left(P_n^{2IP}(\cdot | \mathcal{G}_1^n), \widehat{P}_n^{ILL}(\cdot)\right), \end{aligned}$$

where the first two terms on the right hand side converge in probability to zero by Theorem 2.2(a) and Theorem 1, respectively. We are left to show that  $P_n^{2IP}(\cdot | \mathcal{G}_1^n)$  and  $\widehat{P}_n^{ILL}(\cdot)$  merge in probability. The triangle inequality implies that

$$d_{BL}\left(P_n^{2IP}(\cdot | \mathcal{G}_1^n), \widehat{P}_n^{ILL}(\cdot)\right) \leq d_{BL}\left(P_n^{2IP}(\cdot | \mathcal{G}_1^n), \bar{P}_n^{2IP}(\cdot | \mathcal{G}_1^n)\right) + d_{BL}\left(\bar{P}_n^{2IP}(\cdot | \mathcal{G}_1^n), \widehat{P}_n^{ILL}(\cdot)\right),$$

where  $\bar{P}_n^{2IP}(\cdot | \mathcal{G}_1^n)$  is defined in equation (A.1). In the proof of Theorem 1, we have shown that  $P_n^{2IP}(\cdot | \mathcal{G}_1^n)$  and  $\bar{P}_n^{2IP}(\cdot | \mathcal{G}_1^n)$  merge in probability under Assumption 1. It suffices to show that  $\bar{P}_n^{2IP}(\cdot | \mathcal{G}_1^n)$  and  $\widehat{P}_n^{ILL}(\cdot)$  merge in probability. Write  $\hat{u}'_n \hat{Z}_{\hat{\theta}(\mathbf{X}_n), \hat{\xi}(\mathbf{X}_n)} = w'_n \hat{Z}_{\hat{\theta}(\mathbf{X}_n), \hat{\xi}(\mathbf{X}_n)} + \hat{R}_n^{ILL}$  with  $\hat{R}_n^{ILL} = (\hat{u}_n - w_n)' \hat{Z}_{\hat{\theta}(\mathbf{X}_n), \hat{\xi}(\mathbf{X}_n)}$ . First, we

show  $\hat{R}_n^{ILL} = o_p(1)$ . Take an arbitrary  $\epsilon > 0$ . We obtain

$$\begin{aligned}
\mathbb{P}\left[|\hat{R}_n^{ILL}| \geq \epsilon\right] &\leq \mathbb{P}\left[\left\|\frac{\partial^2 \psi_n(s_0^\circ, \mathbf{X}_n; \ddot{\theta}_n)}{\partial \theta \partial \theta'}\right\| \left\|\hat{\theta}(\mathbf{X}_n) - \theta_0\right\| \left\|\hat{Z}_{\hat{\theta}(\mathbf{X}_n), \hat{\xi}(\mathbf{X}_n)}\right\| \geq \epsilon\right] \\
&\leq \mathbb{P}\left[\left\|\frac{\partial^2 \psi_n(s_0^\circ, \mathbf{X}_n; \ddot{\theta}_n)}{\partial \theta \partial \theta'}\right\| \left\|\hat{\theta}(\mathbf{X}_n) - \theta_0\right\| \left\|\hat{Z}_{\hat{\theta}(\mathbf{X}_n), \hat{\xi}(\mathbf{X}_n)}\right\| \geq \epsilon \cap \ddot{\theta}_n \in \Theta^\star\right] + \mathbb{P}\left[\ddot{\theta}_n \notin \Theta^\star\right] \\
&\leq \mathbb{P}\left[\sup_{\theta \in \Theta^\star} \left\|\frac{\partial^2 \psi_n(s_0^\circ, \mathbf{X}_n; \theta)}{\partial \theta \partial \theta'}\right\| \left\|\hat{\theta}(\mathbf{X}_n) - \theta_0\right\| \left\|\hat{Z}_{\hat{\theta}(\mathbf{X}_n), \hat{\xi}(\mathbf{X}_n)}\right\| \geq \epsilon\right] + \mathbb{P}\left[\ddot{\theta}_n \notin \Theta^\star\right],
\end{aligned}$$

where  $\ddot{\theta}_n$  lies between  $\hat{\theta}(\mathbf{X}_n)$  and  $\theta_0$ . The first term vanishes since  $\sup_{\theta \in \Theta^\star} \left\|\frac{\partial^2 \psi_n(s_0^\circ, \mathbf{X}_n; \theta)}{\partial \theta \partial \theta'}\right\|$  is  $O_p(1)$  by Assumption 1.5,  $\left\|\hat{\theta}(\mathbf{X}_n) - \theta_0\right\| = O_p(m_n^{-1}) = o_p(1)$  by Assumption 1.1 and  $\left\|\hat{Z}_{\hat{\theta}(\mathbf{X}_n), \hat{\xi}(\mathbf{X}_n)}\right\| = O_p(1)$  as  $\hat{Z}_{\hat{\theta}(\mathbf{X}_n), \hat{\xi}(\mathbf{X}_n)} \sim P_{\hat{\theta}(\mathbf{X}_n), \hat{\xi}(\mathbf{X}_n)} \xrightarrow{p} P_{\theta_0, \xi_0}$  by Assumption 4.1. Further, as  $\hat{\theta}(\mathbf{X}_n) \xrightarrow{p} \theta_0 \in \Theta^\star$  and  $\Theta^\star$  is open, we have  $\mathbb{P}[\ddot{\theta}_n \notin \Theta^\star] \rightarrow 0$  and  $\hat{R}_n^{ILL} = o_p(1)$  follows. Moreover,  $w_n$  is  $\mathcal{F}_1^n$ -measurable and  $O_p(1)$  by Assumption 1.6 and  $\int h d\bar{\mathbb{P}}[\hat{Z}_{\hat{\theta}(\mathbf{X}_n), \hat{\xi}(\mathbf{X}_n)} \in \cdot | \mathcal{F}_1^n] = \int h dP_{\hat{\theta}(\mathbf{X}_n), \hat{\xi}(\mathbf{X}_n)} \xrightarrow{p} \int h dP_{\theta_0, \xi_0}$  for each  $h \in BL$ . Replacing  $Z_n$ ,  $R_n$ ,  $w_n$  and  $\mathcal{F}_n$  in Lemma 4 by  $\hat{Z}_{\hat{\theta}(\mathbf{X}_n), \hat{\xi}(\mathbf{X}_n)}$ ,  $\hat{R}_n^{ILL}$ ,  $w_n$  and  $\mathcal{F}_1^n$  implies that  $\widehat{P}_n^{ILL}(\cdot) = \bar{\mathbb{P}}[w'_n \hat{Z}_{\hat{\theta}(\mathbf{X}_n), \hat{\xi}(\mathbf{X}_n)} + \hat{R}_n^{ILL} \in \cdot | \mathcal{F}_1^n]$  and  $\bar{\mathbb{P}}[w'_n Z \in \cdot | \mathcal{F}_1^n] = \bar{\mathbb{P}}[w'_n Z \in \cdot | \mathcal{G}_1^n] = \bar{P}_n^{2IP}(\cdot | \mathcal{G}_1^n)$  merge in probability. Thus,  $\widehat{P}_n^{SPL}(\cdot)$  and  $\widehat{P}_n^{ILL}(\cdot)$  merge in probability. Together with  $\widehat{F}_n^{SPL}(\cdot)$  being stochastically pointwise continuous at  $\gamma_1$  and  $1 - \gamma_2$ , assertion (4.13) follows by Lemma 8, which completes the proof.  $\square$

## B Additional Proofs

### B.1 Proofs of Lemmas

**Proof of Lemma 1.** Consider (i). By Assumption 1.3 one can write  $R_n^{2IP}$  as follows:

$$R_n^{2IP} = \underbrace{m_n(\psi_n(s_0^\circ, \mathbf{X}_n; \theta_0) - \psi_n(S_0, \mathbf{X}_n; \theta_0))}_{=R_{1,n}^{2IP}} + \underbrace{(\hat{\theta}(\mathbf{Y}_n) - \theta_0)' \frac{\partial^2 \psi_n(s_0^\circ, \mathbf{X}_n; \dot{\theta}_n)}{\partial \theta \partial \theta'} m_n(\hat{\theta}(\mathbf{Y}_n) - \theta_0)}_{=R_{2,n}^{2IP}},$$

where  $\dot{\theta}_n$  lies between  $\theta_0$  and  $\hat{\theta}(\mathbf{Y}_n)$ . By Assumption 1.4,  $R_{1,n}^{2IP}$  is  $o_p(1)$ ; hence we are left to show that  $R_{2,n}^{2IP} = o_p(1)$ . Take an arbitrary  $\epsilon > 0$ . We obtain

$$\begin{aligned} \mathbb{P}[|R_{2,n}^{2IP}| \geq \epsilon] &\leq \mathbb{P}\left[\left\|\frac{\partial^2 \psi_n(s_0^\circ, \mathbf{X}_n; \dot{\theta}_n)}{\partial \theta \partial \theta'}\right\| m_n \|\hat{\theta}(\mathbf{Y}_n) - \theta_0\|^2 \geq \epsilon\right] \\ &\leq \mathbb{P}\left[\left\|\frac{\partial^2 \psi_n(s_0^\circ, \mathbf{X}_n; \dot{\theta}_n)}{\partial \theta \partial \theta'}\right\| m_n \|\hat{\theta}(\mathbf{Y}_n) - \theta_0\|^2 \geq \epsilon \cap \dot{\theta}_n \in \Theta^\star\right] + \mathbb{P}[\dot{\theta}_n \notin \Theta^\star] \\ &\leq \mathbb{P}\left[\sup_{\theta \in \Theta^\star} \left\|\frac{\partial^2 \psi_n(s_0^\circ, \mathbf{X}_n; \theta)}{\partial \theta \partial \theta'}\right\| m_n \|\hat{\theta}(\mathbf{Y}_n) - \theta_0\|^2 \geq \epsilon\right] + \mathbb{P}[\dot{\theta}_n \notin \Theta^\star]. \end{aligned}$$

The first term vanishes since  $\sup_{\theta \in \Theta^\star} \left\|\frac{\partial^2 \psi_n(s_0^\circ, \mathbf{X}_n; \theta)}{\partial \theta \partial \theta'}\right\| = O_p(1)$  by Assumption 1.5 and  $m_n \|\hat{\theta}(\mathbf{Y}_n) - \theta_0\|^2 = O_p(m_n^{-1}) = o_p(1)$  by Assumptions 1.1 and 1.2(a). Further, as  $\hat{\theta}(\mathbf{Y}_n) \xrightarrow{p} \theta_0 \in \Theta^\star$  and  $\Theta^\star$  is open, we have  $\mathbb{P}[\dot{\theta}_n \notin \Theta^\star] \rightarrow 0$  and  $R_{2,n}^{2IP} = o_p(1)$  follows.

The proof of (ii) is analogous; by Assumption 2.3 one can express  $R_n^{SPL}$  as follows:

$$R_n^{SPL} = \underbrace{m_n(\psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \theta_0) - \psi_{n_2}(S_{n-n_2}, \mathbf{X}_n^2; \theta_0))}_{=R_{1,n}^{SPL}} + \underbrace{(\hat{\theta}(\mathbf{X}_n^1) - \theta_0)' \frac{\partial^2 \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \ddot{\theta}_n)}{\partial \theta \partial \theta'} m_n(\hat{\theta}(\mathbf{X}_n^1) - \theta_0)}_{=R_{2,n}^{SPL}}$$

with  $R_{1,n}^{SPL} = o_p(1)$  by Assumption 2.4 and  $\ddot{\theta}_n$  lying between  $\theta_0$  and  $\hat{\theta}(\mathbf{X}_n^1)$ . For an arbitrary  $\epsilon > 0$ , we obtain

$$\begin{aligned} \mathbb{P}[|R_{2,n}^{SPL}| \geq \epsilon] &\leq \mathbb{P}\left[\left\|\frac{\partial^2 \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \ddot{\theta}_n)}{\partial \theta \partial \theta'}\right\| m_n \|\hat{\theta}(\mathbf{X}_n^1) - \theta_0\|^2 \geq \epsilon\right] \\ &\leq \mathbb{P}\left[\left\|\frac{\partial^2 \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \ddot{\theta}_n)}{\partial \theta \partial \theta'}\right\| m_n \|\hat{\theta}(\mathbf{X}_n^1) - \theta_0\|^2 \geq \epsilon \cap \ddot{\theta}_n \in \Theta^\star\right] + \mathbb{P}[\ddot{\theta}_n \notin \Theta^\star] \\ &\leq \mathbb{P}\left[\sup_{\theta \in \Theta^\star} \left\|\frac{\partial^2 \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \theta)}{\partial \theta \partial \theta'}\right\| m_n \|\hat{\theta}(\mathbf{X}_n^1) - \theta_0\|^2 \geq \epsilon\right] + \mathbb{P}[\ddot{\theta}_n \notin \Theta^\star]. \end{aligned}$$

The first term vanishes as  $\sup_{\theta \in \Theta^\star} \left\|\frac{\partial^2 \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \theta)}{\partial \theta \partial \theta'}\right\| = O_p(1)$  by Assumption 2.5 and  $m_n \|\hat{\theta}(\mathbf{X}_n^1) - \theta_0\|^2 = O_p(m_n^{-1}) = o_p(1)$  by Assumptions 2.1. Further, as  $\hat{\theta}(\mathbf{X}_n^1) \xrightarrow{p} \theta_0 \in \Theta^\star$  and  $\Theta^\star$  is open, we have  $\mathbb{P}[\ddot{\theta}_n \notin \Theta^\star] \rightarrow 0$  and  $R_{2,n}^{SPL} = o_p(1)$  follows.  $\square$

**Proof of Lemma 2.** Consider (i) and let  $P_{Z_n}^{2IP}$  denote the unconditional distribution of  $Z_n^{2IP}$ . By Assumption 1.2(b), we have for each  $h \in BL$

$$\int h dP_{Z_n}^{2IP}(\cdot | \mathcal{G}_1^n) = \int h dP_{Z_n}^{2IP} \rightarrow \int h dP_Z,$$

where the last assertion comes from Assumptions 1.1 and 1.2(a) and Portmanteau's

Lemma (cf. van der Vaart, 2000; Lem. 2.2). Consider (ii); for each  $h \in BL$  we obtain

$$\int h d(P_{Z_n}^{SPL}(\cdot|\mathcal{G}_{n-n_2+1}^n) - P_Z) = \underbrace{\int h d(P_{Z_n}^{SPL} - P_Z)}_I + \underbrace{\int h d(P_{Z_n}^{SPL}(\cdot|\mathcal{G}_{n-n_2+1}^n) - P_{Z_n}^{SPL})}_{II},$$

where  $I \rightarrow 0$  by Assumption 2.1 and Portmanteau's Lemma and  $II \xrightarrow{p} 0$  by Assumption 2.2.  $\square$

**Proof of Lemma 3.** The proof closely follows Dudley (2002, Thm. 11.3.3). Let  $\epsilon > 0$  and take a compact set  $K \subset \mathbb{R}^r$  such that  $Q(K) > 1 - \epsilon$ . The set of functions  $h \in \mathcal{H}$ , restricted to  $K$ , form a compact set of functions for the supremum norm by the Arzela-Ascoli theorem (cf. Dudley, 2002, Thm. 2.4.7). Thus for some finite  $J = J(\epsilon)$  there are  $h_1, \dots, h_J \in \mathcal{H}$  such that for any  $h \in \mathcal{H}$ , there is a  $j \leq J$  with  $\sup_{y \in K} |h(y) - h_j(y)| < \epsilon$ . Let  $K^\epsilon = \{y \in \mathbb{R}^r : \|x - y\| < \epsilon \text{ for some } x \in K\}$ . One has  $\sup_{x \in K^\epsilon} |h(x) - h_j(x)| < 3\epsilon$ , since if  $y \in K$  and  $\|x - y\| < \epsilon$ , then

$$\begin{aligned} |h(x) - h_j(x)| &\leq |h(x) - h(y)| + |h(y) - h_j(y)| + |h_j(y) - h_j(x)| \\ &\leq \|h\|_{BL}\|x - y\| + \epsilon + \|h_j\|_{BL}\|x - y\| < 3\epsilon. \end{aligned}$$

Let  $g(x) = \max\{0, 1 - \|x - K\|/\epsilon\}$ , where  $\|x - K\| = \inf\{\|x - y\| : y \in K\}$  for all  $x \in \mathbb{R}^r$ . Then  $g \in BL$  and  $I\{x \in K\} \leq g \leq I\{x \in K^\epsilon\}$ , where  $I\{\cdot\}$  denotes the indicator function. It follows that

$$Q_n(\mathbb{R}^r \setminus K^\epsilon) = 1 - Q_n(K^\epsilon) \leq 1 - \int g dQ_n \xrightarrow{p} 1 - \int g dQ \leq 1 - Q(K) < \epsilon$$

or equivalently  $\mathbb{P}[Q_n(\mathbb{R}^r \setminus K^\epsilon) \geq \epsilon] \rightarrow 0$ . Thus, for each  $h \in \mathcal{H}$  and  $h_j$  as above

$$\begin{aligned}
\sup_{h \in \mathcal{H}} \left| \int h d(Q_n - Q) \right| &\leq \sup_{h \in \mathcal{H}} \int |h - h_j| d(Q_n + Q) + \sup_{h \in \mathcal{H}} \left| \int h_j d(Q_n - Q) \right| \\
&\leq 2(Q_n + Q)(\mathbb{R}^r \setminus K^\epsilon) + 6\epsilon + \max_{1 \leq j \leq J} \left| \int h_j d(Q_n - Q) \right| \\
&\leq 8\epsilon + 2Q_n(\mathbb{R}^r \setminus K^\epsilon) + \max_{1 \leq j \leq J} \left| \int h_j d(Q_n - Q) \right|.
\end{aligned}$$

Hence,

$$\begin{aligned}
&\mathbb{P} \left[ \sup_{h \in \mathcal{H}} \left| \int h d(Q_n - Q) \right| \geq 11\epsilon \right] \\
&\leq \mathbb{P} \left[ 2Q_n(\mathbb{R}^r \setminus K^\epsilon) + \max_{1 \leq j \leq J} \left| \int h_j d(Q_n - Q) \right| \geq 3\epsilon \right] \\
&\leq \mathbb{P} \left[ Q_n(\mathbb{R}^r \setminus K^\epsilon) \geq \epsilon \right] + \mathbb{P} \left[ \max_{1 \leq j \leq J} \left| \int h_j d(Q_n - Q) \right| \geq \epsilon \right] \\
&\leq \mathbb{P} \left[ Q_n(\mathbb{R}^r \setminus K^\epsilon) \geq \epsilon \right] + \sum_{j=1}^J \mathbb{P} \left[ \left| \int h_j d(Q_n - Q) \right| \geq \epsilon \right],
\end{aligned}$$

where the last two terms are converging to 0 for finite  $J$ . □

**Proof of Lemma 5.** Take  $\epsilon > 0$  and let  $F$  and  $G$  be cdfs on  $\mathbb{R}$  with  $G(\tau - \epsilon) - \epsilon \leq$



$F(\tau) \leq G(\tau + \epsilon) + \epsilon$  for all  $\tau \in \mathbb{R}$ . Fixing  $u \in (\epsilon, 1 - \epsilon)$ , we obtain

$$\inf \{ \tau \in \mathbb{R} : F(\tau) \geq u + \epsilon \} + \epsilon \tag{B.1}$$

$$\begin{aligned} &\geq \inf \{ \tau \in \mathbb{R} : G(\tau + \epsilon) + \epsilon \geq u + \epsilon \} + \epsilon \\ &= \inf \{ \tau \in \mathbb{R} : G(\tau + \epsilon) \geq u \} + \epsilon \\ &= \inf \{ \tau + \epsilon, \tau \in \mathbb{R} : G(\tau + \epsilon) \geq u \} \\ &= \inf \{ \tau \in \mathbb{R} : G(\tau) \geq u \} \end{aligned} \tag{B.2}$$

$$\begin{aligned} &= \inf \{ \tau + \epsilon, \tau \in \mathbb{R} : G(\tau) \geq u \} - \epsilon \\ &= \inf \{ \tau \in \mathbb{R} : G(\tau - \epsilon) \geq u \} - \epsilon \\ &= \inf \{ \tau \in \mathbb{R} : G(\tau - \epsilon) - \epsilon \geq u - \epsilon \} - \epsilon \\ &\geq \inf \{ \tau \in \mathbb{R} : F(\tau) \geq u - \epsilon \} - \epsilon. \end{aligned} \tag{B.3}$$

Identifying (B.1), (B.2) and (B.3) as  $F^{-1}(u + \epsilon) + \epsilon$ ,  $G^{-1}(u)$  and  $F^{-1}(u - \epsilon) - \epsilon$ , respectively, completes the proof.  $\square$

**Proof of Lemma 6.** Let  $\epsilon, \eta > 0$ . As  $G_n$  is stochastically uniformly equicontinuous, there exists a  $\delta > 0$  and an  $N_1 \in \mathbb{N}$  such that  $\mathbb{P} \left[ \sup_{\tau \in \mathbb{R}} \sup_{\tau' \in \mathbb{R} : |\tau - \tau'| < \delta} |G_n(\tau') - G_n(\tau)| > \epsilon \right] < \eta$  for all  $n \geq N_1$ . Take  $\varkappa = \min(\delta/2, \epsilon)$ . As  $d_L(F_n, G_n) \xrightarrow{p} 0$  as  $n \rightarrow \infty$ , there

exists an  $N_2$  such that  $\mathbb{P}[d_L(F_n, G_n) > \varkappa] < \eta$  for all  $n \geq N_2$ . Let  $N = \max(N_1, N_2)$ .

$$\begin{aligned}
& \mathbb{P}\left[\sup_{\tau \in \mathbb{R}} |F_n(\tau) - G_n(\tau)| > 2\epsilon\right] \\
& \leq \mathbb{P}\left[\sup_{\tau \in \mathbb{R}} |F_n(\tau) - G_n(\tau)| > 2\epsilon \cap d_L(F_n, G_n) \leq \varkappa\right] + \mathbb{P}[d_L(F_n, G_n) > \varkappa] \\
& \leq \mathbb{P}\left[\varkappa + \sup_{\tau \in \mathbb{R}} |G_n(\tau \pm \varkappa) - G_n(\tau)| > 2\epsilon\right] + \mathbb{P}[d_L(F_n, G_n) > \varkappa] \\
& \leq \mathbb{P}\left[\sup_{\tau \in \mathbb{R}} \sup_{\tau' \in \mathbb{R}: |\tau - \tau'| < \delta} |G_n(\tau') - G_n(\tau)| > \epsilon\right] + \mathbb{P}[d_L(F_n, G_n) > \varkappa] < 2\eta
\end{aligned}$$

for all  $n \geq N$ . Since the choice of  $\epsilon$  and  $\eta$  was arbitrary, the desired result follows.  $\square$

**Proof of Lemma 8.** Let  $\epsilon, \eta > 0$  and set  $\bar{\epsilon} = \min\{\epsilon, u, 1 - u\}/2$ . Since  $G_n^{-1}$  is point-wise equicontinuous at  $u$ , there exist a  $\delta > 0$  and an  $N_1$  such that  $\mathbb{P}\left[\sup_{|u-v| < \delta} |G_n^{-1}(v) - G_n^{-1}(u)| > \bar{\epsilon}\right] < \eta$  for all  $n \geq N_1$ . Take  $\varkappa = \min\{\delta/2, \bar{\epsilon}\}$ . As  $d_L(F_n, G_n) \xrightarrow{p} 0$  as  $n \rightarrow \infty$ , there exists an  $N_2$  such that  $\mathbb{P}[d_L(F_n, G_n) > \varkappa] < \eta$  for all  $n \geq N_2$ .

$$\begin{aligned}
& \mathbb{P}\left[|F_n^{-1}(u) - G_n^{-1}(u)| > 2\epsilon\right] \leq \mathbb{P}\left[|F_n^{-1}(u) - G_n^{-1}(u)| > 2\bar{\epsilon}\right] \\
& \leq \mathbb{P}\left[|F_n^{-1}(u) - G_n^{-1}(u)| > 2\bar{\epsilon} \cap d_L(F_n, G_n) \leq \varkappa\right] + \mathbb{P}[d_L(F_n, G_n) > \varkappa] \\
& \leq \mathbb{P}\left[\varkappa + |G_n^{-1}(u \pm \varkappa) - G_n^{-1}(u)| > 2\bar{\epsilon}\right] + \mathbb{P}[d_L(F_n, G_n) > \varkappa] \\
& \leq \mathbb{P}\left[\sup_{|u-v| < \delta} |G_n^{-1}(v) - G_n^{-1}(u)| > \bar{\epsilon}\right] + \mathbb{P}[d_L(F_n, G_n) > \varkappa] < 2\eta
\end{aligned}$$

for all  $n \geq N = \max(N_1, N_2)$ , where the third inequality follows from Lemma 5 and  $u \in (\bar{\epsilon}, 1 - \bar{\epsilon}) \subseteq (\varkappa, 1 - \varkappa)$ . As  $\epsilon$  and  $\eta$  were arbitrarily chosen, this completes the proof.  $\square$

## B.2 Proofs of Corollaries

**Proof of Corollary 1.** Statement 1(a) follows immediately from Theorem 2.1(a) and  $\widehat{P}_n^{2IP}(\cdot)$  equals to  $N(0, \hat{v}_n^{2IP})$ . Regarding claim 1(b), it is sufficient to show that  $1/\hat{v}_n^{2IP} = O_p(1)$  implies that  $\widehat{F}_n^{2IP}(\cdot) = \Phi(\cdot / \sqrt{\hat{v}_n^{2IP}})$  is stochastically uniformly equicontinuous by Theorem 2.1(b). Since  $1/\hat{v}_n^{2IP} = O_p(1)$  by assumption, we have for all  $\varkappa > 0$ , there exist  $K = K(\varkappa)$  and  $N = N(\varkappa)$  such that  $\mathbb{P}[1/\hat{v}_n^{2IP} > K] < \varkappa$  for all  $n > N$ . Let  $\phi$  denote the standard normal density. Taking  $\delta = \frac{\epsilon}{\phi(0)\sqrt{K}}$ , we obtain

$$\begin{aligned} & \mathbb{P}\left[\sup_{\tau \in \mathbb{R}} \sup_{\tau': |\tau - \tau'| < \delta} |\Phi(\tau'/\sqrt{\hat{v}_n^{2IP}}) - \Phi(\tau/\sqrt{\hat{v}_n^{2IP}})| > \epsilon\right] \\ & \leq \mathbb{P}\left[\phi(0)\delta/\sqrt{\hat{v}_n^{2IP}} > \epsilon\right] = \mathbb{P}\left[1/\hat{v}_n^{2IP} > K\right] < \varkappa \end{aligned}$$

for all  $n > N$  such that the stochastic uniform equicontinuity condition holds.

Statement 2(a) follows from Theorem 2.2(a) and  $\widehat{P}_n^{SPL}(\cdot)$  equals to  $N(0, \hat{v}_n^{SPL})$ . Claim 2(b) is proven analogously to the claim of 1(b) replacing  $\hat{v}_n^{2IP}$  and  $\widehat{F}_n^{2IP}(\cdot)$  with  $\hat{v}_n^{SPL}$  and  $\widehat{F}_n^{SPL}(\cdot)$ , respectively.  $\square$

**Proof of Corollary 2.** In the proof of Theorem 3 we have shown that  $\widehat{P}_n^{SPL}(\cdot)$  and  $\widehat{P}_n^{ILL}(\cdot)$  merge in probability, which simplify to  $N(0, \hat{v}_n^{SPL})$  and  $N(0, \hat{v}_n^{ILL})$ , respectively, under Assumption 5. It remains to show that  $\widehat{F}_n^{SPL}{}^{-1}(u) = \sqrt{\hat{v}_n^{SPL}}\Phi^{-1}(u)$  is stochastically pointwise equicontinuous at  $u = \gamma_1, 1 - \gamma_2$ . First, we show that  $\hat{v}_n^{SPL} = O_p(1)$ . The triangle inequality implies  $\hat{v}_n^{SPL} \leq v_n^{SPL} + |\hat{v}_n^{SPL} - v_n^{SPL}|$ , where  $v_n^{SPL} = \frac{\partial \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \theta_0)}{\partial \theta'} \Upsilon_0 \frac{\partial \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \theta_0)}{\partial \theta}$  is  $O_p(1)$  by Assumption 2.6. Moreover, for

an arbitrary  $\varepsilon > 0$ , we have

$$\begin{aligned}
& \mathbb{P} \left[ \left\| \frac{\partial \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \theta_0)}{\partial \theta} - \frac{\partial \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \hat{\theta}(\mathbf{X}_n^1))}{\partial \theta} \right\| \geq \epsilon \right] \\
& \leq \mathbb{P} \left[ \left\| \frac{\partial^2 \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \tilde{\theta}_n)}{\partial \theta \partial \theta'} \right\| \left\| \hat{\theta}(\mathbf{X}_n^1) - \theta_0 \right\| \geq \epsilon \cap \tilde{\theta}_n \in \Theta^\star \right] + \mathbb{P}[\tilde{\theta}_n \notin \Theta^\star] \\
& \leq \mathbb{P} \left[ \sup_{\theta \in \Theta^\star} \left\| \frac{\partial^2 \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \theta)}{\partial \theta \partial \theta'} \right\| \left\| \hat{\theta}(\mathbf{X}_n^1) - \theta_0 \right\| \geq \epsilon \right] + \mathbb{P}[\tilde{\theta}_n \notin \Theta^\star],
\end{aligned}$$

where  $\tilde{\theta}_n$  lies between  $\hat{\theta}(\mathbf{X}_n^1)$  and  $\theta_0$ . The first term vanishes as  $\sup_{\theta \in \Theta^\star} \left\| \frac{\partial^2 \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \theta)}{\partial \theta \partial \theta'} \right\| = O_p(1)$  by Assumption 2.5 and  $\left\| \hat{\theta}(\mathbf{X}_n^1) - \theta_0 \right\| = O_p(m_n^{-1}) = o_p(1)$  by Assumptions 2.1. Further, since  $\hat{\theta}(\mathbf{X}_n^1) \xrightarrow{p} \theta_0 \in \Theta^\star$  and  $\Theta^\star$  is open, we have  $\mathbb{P}[\tilde{\theta}_n \notin \Theta^\star] \rightarrow 0$  and  $\left\| \frac{\partial \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \theta_0)}{\partial \theta} - \frac{\partial \psi_{n_2}(s_{n-n_2}^\circ, \mathbf{X}_n^2; \hat{\theta}(\mathbf{X}_n^1))}{\partial \theta} \right\| = o_p(1)$  follows. Together with  $\hat{\Upsilon}(\mathbf{X}_n^1) \xrightarrow{p} \Upsilon_0$ , it implies  $|\hat{v}_n^{SPL} - v_n^{SPL}| = o_p(1)$  and hence  $\hat{v}_n^{SPL} = O_p(1)$ . Next, we show that the stochastic pointwise equicontinuity condition is satisfied. For  $K > 0$ , we get

$$\begin{aligned}
& \mathbb{P} \left[ \sqrt{\hat{v}_n^{SPL}} \sup_{v: |u-v| < \delta} |\Phi^{-1}(u) - \Phi^{-1}(v)| > \epsilon \right] \\
& \leq \mathbb{P} \left[ \sqrt{K} \sup_{v: |u-v| < \delta} |\Phi^{-1}(u) - \Phi^{-1}(v)| > \epsilon \right] + \mathbb{P}[\hat{v}_n^{SPL} > K].
\end{aligned}$$

$K$  can be chosen such that the last term is arbitrary small for large  $n$  as  $\hat{v}_n^{SPL} = O_p(1)$ .

Given  $K$ , the first term is 0 by the choice of  $\delta$  and continuity of  $\Phi^{-1}$ .  $\square$

## References

- Akaike, H. (1969). Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics*, 21(1):243–247.
- Barndorff-Nielsen, O. E. and Cox, D. R. (1996). Prediction and asymptotics. *Bernoulli*, 2(4):319–340.
- Belyaev, Y. and Sjöstedt-De Luna, S. (2000). Weakly approaching sequences of random distributions. *Journal of Applied Probability*, 37(3):807–822.

- Beutner, E., Heinemann, A., and Smeekes, S. (2017). Technical report on “A justification of conditional confidence intervals”. Technical report, Maastricht University, <http://researchers-sbe.unimaas.nl/stephansmeekes/research/>.
- Billingsley, P. (1986). *Probability and Measure*. John Wiley & Sons, New York.
- Blasques, F., Koopman, S. J., Lasak, K., and Lucas, A. (2016). In-sample confidence bands and out-of-sample forecast bands for time-varying parameters in observation-driven models. *International Journal of Forecasting*, 32(3):875–887.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.
- Boussama, F., Fuchs, F., and Stelzer, R. (2011). Stationarity and geometric ergodicity of BEKK multivariate GARCH models. *Stochastic Processes and their Applications*, 121(10):2331–2360.
- Castillo, I. and Rousseau, J. (2015). A Bernstein–von Mises theorem for smooth functionals in semiparametric models. *Annals of Statistics*, 43(6):2353–2383.
- Cavaliere, G., Georgiev, I., and Taylor, A. M. R. (2013). Wild bootstrap of the sample mean in the infinite variance case. *Econometric Reviews*, 32(2):204–219.
- Creal, D., Koopman, S. J., and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795.
- D’Aristotile, A., Diaconis, P., and Freedman, D. (1988). On merging of probabilities. *Sankhyā: The Indian Journal of Statistics, Series A*, 50(3):363–380.
- Davydov, Y. and Rotar, V. (2009). On asymptotic proximity of distributions. *Journal of Theoretical Probability*, 22(1):82–98.
- Doukhan, P. (1994). *Mixing: Properties and Examples*. Springer, New York.
- Dudley, R. M. (2002). *Real Analysis and Probability*. Cambridge University Press, Cambridge.
- Dufour, J.-M. and Taamouti, A. (2010). Short and long run causality measures: theory and inference. *Journal of Econometrics*, 154(1):42–58.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007.
- Engle, R. F. and Russell, J. R. (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, 66(5):1127–1162.
- Francq, C. and Zakoïan, J.-M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli*, 10(4):605–637.

- Francq, C. and Zakoïan, J.-M. (2015). Risk-parameter estimation in volatility models. *Journal of Econometrics*, 184(1):158–173.
- Gospodinov, N. (2002). Median unbiased forecasts for highly persistent autoregressive processes. *Journal of Econometrics*, 111(1):85–101.
- Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, London.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, Princeton.
- Hansen, B. E. (2006). Interval forecasts and parameter uncertainty. *Journal of Econometrics*, 135(1):377–398.
- Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. Cambridge University Press, New York.
- Huber, P. J. (2009). *Robust Statistics*. John Wiley & Sons, New York, 2. edition.
- Kabaila, P. (1999). The relevance property for prediction intervals. *Journal of Time Series Analysis*, 20(6):655–662.
- Kabaila, P. and He, Z. (2004). The adjustment of prediction intervals to account for errors in parameter estimation. *Journal of Time Series Analysis*, 25(3):351–358.
- Kabaila, P. and Syuhada, K. (2008). Improved prediction limits for  $AR(p)$  and  $ARCH(p)$  processes. *Journal of Time Series Analysis*, 29(2):213–223.
- Kabaila, P. and Syuhada, K. (2010). The asymptotic efficiency of improved prediction intervals. *Statistics & Probability Letters*, 80(17):1348–1353.
- Klenke, A. (2014). *Probability Theory: A Comprehensive Course*. Springer, Berlin, 2 edition.
- Kreiss, J.-P. (2016). Discussion: bootstrap prediction intervals for linear, nonlinear and nonparametric autoregressions. *Journal of Statistical Planning and Inference*, 177:28–30.
- Kunitomo, N. and Yamamoto, T. (1985). Properties of predictors in misspecified autoregressive time series models. *Journal of the American Statistical Association*, 80(392):941–950.
- Lewis, R. and Reinsel, G. C. (1985). Prediction of multivariate time series by autoregressive model fitting. *Journal of Multivariate Analysis*, 16(3):393–411.
- Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer, Berlin.

- Pan, L. and Politis, D. N. (2016a). Bootstrap prediction intervals for linear, nonlinear and nonparametric autoregressions. *Journal of Statistical Planning and Inference*, 177:1–27.
- Pan, L. and Politis, D. N. (2016b). Bootstrap prediction intervals for Markov processes. *Computational Statistics & Data Analysis*, 100:467–494.
- Pascual, L., Romo, J., and Ruiz, E. (2004). Bootstrap predictive inference for ARIMA processes. *Journal of Time Series Analysis*, 25(4):449–465.
- Pascual, L., Romo, J., and Ruiz, E. (2006). Bootstrap prediction for returns and volatilities in GARCH models. *Computational Statistics & Data Analysis*, 50(9):2293–2312.
- Pesaran, M. H. (2015). *Time Series and Panel Data Econometrics*. Oxford University Press, Oxford.
- Phillips, P. C. B. (1979). The sampling distribution of forecasts from a first-order autoregression. *Journal of Econometrics*, 9(3):241–261.
- Samaranayake, V. A. and Hasza, D. P. (1988). Properties of predictors for multivariate autoregressive models with estimated parameters. *Journal of Time Series Analysis*, 9(4):361–383.
- Schmidt, P. (1977). Some small sample evidence on the distribution of dynamic simulation forecasts. *Econometrica*, 45(4):997–1005.
- van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press, Cambridge.
- Vidoni, P. (2004). Improved prediction intervals for stochastic process models. *Journal of Time Series Analysis*, 25(1):137–154.
- Vidoni, P. (2009a). Improved prediction intervals and distribution functions. *Scandinavian Journal of Statistics*, 36(4):735–748.
- Vidoni, P. (2009b). A simple procedure for computing improved prediction intervals for autoregressive models. *Journal of Time Series Analysis*, 30(6):577–590.
- Vidoni, P. (2017). Improved multivariate prediction regions for Markov process models. *Statistical Methods & Applications*, 26(1):1–18.
- Xiong, S. and Li, G. (2008). Some results on the convergence of conditional distributions. *Statistics & Probability Letters*, 78(18):3249–3253.
- Zakoïan, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, 18(5):931–955.